

Master Thesis

Deep Learning-Based Super-Resolution and De-Noising for XMM-Newton EPIC-pn

S.F. Sweere

Thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science of Artificial Intelligence
at the Department of Data Science and Knowledge Engineering
of the Maastricht University

Thesis Committee:

Dr. A. Briassouli
Dr. D.H. Cámpora Pérez

Maastricht University
Faculty of Science and Engineering
Department of Data Science and Knowledge Engineering

March 22, 2022

Abstract

The field of AI image enhancement has been rapidly evolving over the last few years and is able to produce impressive results on non-astronomical images. In this thesis we present the first application of Machine Learning based super-resolution (SR) and de-noising (DN) to enhance X-ray images from *XMM-Newton*. This telescope has over the last two decades advanced our understanding of black holes, the formation of galaxies, and many other phenomena. We propose *XMM-SuperRes* and *XMM-DeNoise*, deep-learning-based models that can generate enhanced SR and DN images from real observations. The models are trained on realistic *XMM-Newton* simulations, developed for this thesis, such that *XMM-SuperRes* is able to enhance images to twice the spatial resolution with improved noise characteristics. The *XMM-DeNoise* enhances images comparable to a $2.5 \times$ increase in exposure time and improves the peak-signal-to-noise ratio by 8.2% and reduces the L1 by 38% when tested on real observations. These enhanced images allow identification of features that are otherwise hard or impossible to perceive in the original images. We demonstrate the feasibility of using our deep-learning models to enhance *XMM-Newton* X-ray images to increase their scientific value in a way that could benefit the legacy of the *XMM-Newton* archives.

Contents

1	Introduction	4
1.1	Research Questions	6
1.2	Contributions	6
1.3	Thesis Structure	7
2	Preliminaries	8
2.1	XMM-Newton	8
2.1.1	X-ray Astronomy	8
2.1.2	XMM-Newton Summary	9
2.1.3	Telescope Lens Properties	10
2.1.4	Telescope Sensor Properties	12
2.1.5	Background Noise	15
2.1.6	Image Generation	16
2.2	Super-Resolution and De-Noising Methods	17
2.2.1	Super-Resolution Methods	17
2.2.2	De-Noising Methods	18
3	Data	19
3.1	Real XMM-Newton Dataset	19
3.2	Simulated XMM-Newton Dataset	20
3.2.1	Our Approach	22
3.3	Simulation Software Setup	22
3.3.1	Simulated Telescope Lens Configuration	23
3.3.2	Simulated EPIC-pn Sensor Configuration	23
3.4	Simulation Input	25
3.4.1	Extended Source Component	25
3.4.2	AGN Component	26
3.4.3	Background	27
3.4.4	Simulated XMM Training Data	28
3.4.5	Train, Validation and Test split	29
4	Experimental Methodology	30
4.1	Data Prepossessing	30
4.1.1	Data Scaling	30

1	Introduction	1
2	Methodology	16
3	Model Architecture	16
4	Training and Evaluation	16
4.1	Training Data Generation	16
4.2	Super-Resolution and De-Noising Model	32
4.2.1	Model Architectures	33
4.2.2	Loss Functions	34
4.2.3	Image Quality Metrics	35
4.3	Hyper-Parameter Tuning	36
4.3.1	Model Hyper-Parameter Tuning	37
4.3.2	Data Hyper-Parameter Tuning	39
4.4	Final Model Training	42
5	Results	44
5.1	Super-Resolution Model	44
5.1.1	Visual Analysis	45
5.1.2	Brightness Analysis	46
5.1.3	XMM-SuperRes Model	49
5.1.4	Chandra Comparison	49
5.2	De-Noising Model	53
5.2.1	Wavelet Comparison	54
6	Discussion	55
6.1	Research Questions	55
6.2	Limitations and Future Work	58
6.2.1	Reliability	58
6.2.2	Bright Sources	58
6.2.3	Usable Domain	59
6.2.4	Simulation Limitations	59
6.2.5	AGN Deblending	59
7	Conclusion	60
8	Acknowledgements	62
A	Appendix	69
A.1	Real XMM Dataset Generation	69
A.2	Simulation Setup Details	70
A.2.1	Boresight Determination	70
A.2.2	PSF	70
A.3	Image Quality Metrics	71
A.3.1	L1	71
A.3.2	Poisson	71
A.3.3	PSNR	71
A.3.4	SSIM	72
A.3.5	MS_SSIM	72
A.3.6	FSIM and HaarPSI	72
A.4	Loss Function Normalization	73
A.5	IllustrisTNG Simput	74
A.6	Simulated XMM-Newton Images	75

A.7	Full-Frame Super-resolution and De-noising	76
A.8	Full-Frame De-noised and Wavelet	77
A.9	Bright Sources Limitation	78

Chapter 1

Introduction

Over the last two decades, the *XMM-Newton* [1] X-ray space observatory has been continuously advancing our understanding of the cosmos through detailed observations of black holes, the formation of galaxies and many other phenomena in our X-ray sky [2, 3].

XMM-Newton, a satellite in orbit around the Earth, has 3 X-ray telescopes on-board equipped with a set of CCD detectors. These European Photon Imaging Cameras (EPIC), consist of two MOS-CCD arrays [4] and a pn-CCD [5]. EPIC-pn has an effective area ~ 3 times that of a single MOS. However, its larger point-spread-function (PSF) and larger pixel size result in a poorer spatial resolution in comparison to MOS. NASA's *Chandra* X-ray telescope [6] has a spatial resolution superior to both the pn and MOS detectors, but with a much smaller effective area and hence spectral resolution. It is desirable to have both good spectral and spatial resolutions in our observations. Longer exposures allow us to collect more photons and hence pick up fainter sources. However, the noise will also increase, thus there is a need for better sensitivity in order to detect these fainter sources. With the current limitations of existing X-ray observatories, it is not uncommon for studies to make use of the high sensitivity of *XMM-Newton* for detections and later follow-up with high-resolution observations using *Chandra*.

Unfortunately, in astronomy, telescopes are generally oversubscribed. There is much competition for the valuable observation time. While ground-based observations may be cheaper, the signal-to-noise ratio (SNR) tends to be lower in space-based observatories since the Earth's atmosphere absorbs most of the X-rays. Image enhancement through noise level reduction also referred to as de-noising, is a popular way to improve the SNR [7].

In X-ray observations, photon counts are subject to a Poissonian shot noise¹, that is dependent on the count rate itself. Therefore, the SNR is smaller in low count rate areas, limiting the detection of faint sources. Binning of X-ray photons is one way to increase the total SNR, albeit at the cost of reducing

¹Noise that originates from discrete nature of converting photons into an electric charge which follows a Poisson distribution.

spatial resolution. [8] uses an adaptive binning method on Chandra observations of the Perseus cluster to reveal structure in the central region. [9] introduced a multi-scale wavelet transform approach to de-noise MOS1 and MOS2 images and were able to successfully recover the total flux and signal shape of toy-model sources, demonstrating that de-noising methods like these can be used to provide more accurate brightness maps from images.

In addition to the noise, an instrument’s point spread function (PSF) can smear out point sources. Sources that are smaller than the PSF can be mistaken for a larger source, and on occasion, two or more sources close together on the sky, can be mistaken as a single extended source. Resolving and deblending such sources can be achieved with super-resolution. Super-resolution (SR) describes a class of methods that can upscale images or video from lower resolutions to higher ones. Such methods have been successfully demonstrated on astronomical imaging, e.g. [10, 11, 12]. Many methods for SR exist [13, 14].

Traditionally, SR uses interpolation methods such as bilinear and nearest neighbour interpolation for upscaling. However, these methods often introduce side effects such as noise amplification and blurring. Furthermore, super-resolution on X-ray images imposes additional challenges since X-ray images are typically sparse, and the data are Poisson distributed. Nevertheless, [15] demonstrate using a direct demodulation (DD) method, the spatial resolution of XMM EPIC images can be improved by a factor of 5 while adhering to the requirements for spectral studies.

In recent years significant progress has been made in the field of de-noising and super-resolution using machine learning methods [16, 17, 18, 19, 20, 21, 22, 23]. In these learning-based approaches, a network is trained with data to learn the mapping between an image and an enhanced image, where the enhanced image in our case is a higher SNR image and/or a higher resolution image. These researches have primarily made use of images captured on Earth or computer-simulated images such as games and art.

Super-resolution and de-noising is fundamentally an ill-posed problem since given a noisy/low-resolution input image, there are an infinite number of possible enhanced (high resolution) images that it could correspond to. The noisy input image inherently does not contain all the information of an enhanced image.

These models primarily make use of Fully Convolutional Networks (FCN), trained using a relevant quantitative metric used as a loss function. Similar to traditional convolutional neural networks [24], FCNs comprise of convolutional, pooling and layers, however they do not have dense layers and their output size are typically the same or larger than the input. For this reason, FCNs are often used for computer vision tasks such as semantic segmentation, de-noising and super-resolution [23, 22, 20]. Images generated this way tend to lack clarity as they often minimise a *simple* loss function such as the mean absolute error (L1), that favors predicting the average over all plausible enhanced images. This leads to fewer finer detail in the generated images. To address this, more recent approaches make use of more complex loss functions. The perceptual loss function [25, 17] incorporates style transfer through pre-training on a target dataset with a particular style or content. Generative Adversarial Networks (GANs)

[26] use two competing models - a generator to produce enhanced images from a given input image and a discriminator to differentiate between the real and generated images. Such networks make use of an adversarial loss [19] to obtain photorealistic images.

In astronomy, these methods have been used to improve observations. [27] showed that using a GAN, they were able to recover features from artificially degraded optical observations. [7] used CNN's and perceptual loss to de-noising Hubble Space Telescope to improve the signal-to-noise ratio by a factor of 1.3-1.5, and [28] demonstrates super-resolution in the in the sub-millimetre spectrum. With a model that generates higher resolution JCMT SCUBA-like observations given a *Herschel* SPIRE observation. However, at the time of writing no published research has been done on applying these techniques on *XMM-Newton* observations.

In this thesis we aim to apply these ideas and develop deep learning-based methods for super-resolution and de-noising of images from *XMM-Newton* to increase their scientific value. The *XMM-Newton* Science Archive contains observations spanning over 20 years. There is ample data to satiate the training of a machine learning model. Improving the quality of this existing data is of great interest to the astronomical community and the lasting legacy of *XMM-Newton*.

1.1 Research Questions

The in-depth research for this thesis started of by focusing on the following research questions:

1. How can we use GAN models to increase the resolution of XMM- Newton X-ray images and increase their scientific value?
2. How can we use U-net CNN-based models to de-noise XMM-Newton X-ray images and increase their scientific value?
3. How does the scientific reliability of the improved images compare with published results using conventional, non-ML methods?
4. If the scientific validation from point 3 is successful, can we obtain new results on potentially interesting targets and publish them?

1.2 Contributions

The key contributions of this thesis can be summarized as:

- We propose a SR and DN model named *XMM-SuperRes*. This model generates enhanced images from real *XMM-Newton* observations with twice the spatial resolution and an improved PSNR.
- We propose a DN model named *XMM-DeNoise* based on the *XMM-SuperRes* model, which aims to generate images with 2.5 times higher exposure without increasing the resolution.

- In order to train these models, we created a custom *XMM-Newton* EPIC-pn simulator to generate both realistic *XMM-Newton* images and improved *XMM-Newton* images with twice the spatial resolution and higher exposure times.
- We highlight challenges that arise when training SR and DN models using astronomical X-ray images and propose using a combined loss function using both PSNR and MS_SSIM and data scaling to solve these challenges.
- We analyse and validate the performance of the *XMM-SuperRes* model using image quality metrics, cutout plots and comparison to NASA’s higher resolution *Chandra* X-ray telescope.
- We research the training of the DN model on both simulated and real data and show that the best performance is gained when the model is trained on simulated data first and then fine-tuned on real data.
- We analyse and validate the performance of the *XMM-DeNoise* model using image quality metrics, comparing it to ground-truth data and to non-ML wavelet transformed images.

1.3 Thesis Structure

In chapter 2 we discuss preliminary information to this thesis, including the *XMM-Newton* X-ray telescope, X-ray astronomy, super-resolution and de-noising methods. In chapter 3 we introduce the simulations that are used to train and validate our method, we describe the pre-processing techniques applied to this data and additionally present the real observational data that our method is implemented on. After declaring our models and model choices in chapter 4, we present our results on the simulated and real observations in chapter 5. We discuss and present our conclusions in chapter 6 and chapter 7 respectively.

Chapter 2

Preliminaries



Figure 2.1: Artist impression of the XMM-Newton X-ray space observatory.¹

This chapter discusses some key concepts that are required to understand this research. First, an introduction to the *XMM-Newton* telescope and X-ray astronomy is given (section 2.1). Next, traditional super-resolution and de-noising methods are discussed (section 2.2).

2.1 XMM-Newton

In this research we are focusing on data from the *XMM-Newton* (Figure 2.1). The *XMM-Newton* is an X-ray telescope in orbit around the Earth. In this section, we will shortly introduce X-ray astronomy, why we are specifically interested in the *XMM-Newton* telescope and how the telescope operates. Finally, we will discuss a few concepts used in X-ray astronomy that are helpful for understanding this thesis.

2.1.1 X-ray Astronomy

X-ray astronomy studies X-ray emissions from celestial objects. X-rays are a part of the electromagnetic spectrum. They have a short wavelength, 10^{-9} to

¹Source: NASA <https://science.nasa.gov/get-involved/toolkits/spacecraft-icons>.

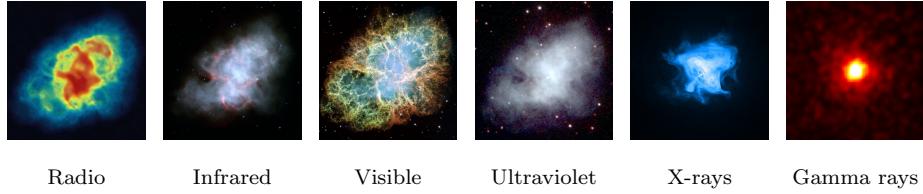


Figure 2.2: Crab Nebula observed at different wavelengths.²

10^{-12} meters and a high energy ± 0.1 to 100 keV (kilo electron volt). Since X-rays have such short wavelengths, they are usually expressed by their energy. X-rays observed in astronomy are created by extremely hot gasses with temperatures ranging from a million kelvin to hundreds of millions of kelvin. Astrophysical objects that emit X-rays include galaxy clusters, black holes, supernova remnants, and stars. When observing celestial objects at different wavelengths, different properties become visible. For example, in Figure 2.2 the Crab Nebula, a supernova remnant, is shown at different wavelengths.

The atmosphere of the Earth absorbs practically all X-rays. Therefore if we want to do astronomical observations in the X-ray spectrum, we need to put our telescope outside our atmosphere.

2.1.2 XMM-Newton Summary

The *XMM-Newton* space observatory was launched on December 10, 1999, by the European Space Agency (ESA). [1] The 4-tonne, 10-meter long spacecraft has a highly eccentric 48h orbit, with a perigee (closest to Earth) of 7000 km and an apogee (furthest from Earth) of 114000 km, almost one-third the distance to the moon. The reason for this far out orbit is twofold. Firstly, the *XMM-Newton* instruments need to be operated outside the radiation belts surrounding the Earth. These radiation belts contain high-energy particles that can damage the instruments and cause false readings. Secondly, the eccentric orbit makes it possible to have longer observation periods without being interrupted by the Earth's shadow.

The *XMM-Newton* contains multiple instruments and has three primary telescopes. In this research, we will be focusing on data from the EPIC-pn (European Photon Imaging Camera pn-CCD) imaging detector. [29] We will therefore only introduce concepts regarding this sensor.

The imaging system consists of two main elements. The lens, focusing the X-ray photons (subsection 2.1.3), and the sensor, recording X-rays that hit it (subsection 2.1.4). Finally, these recordings are converted to images (subsection 2.1.6).

²Source: NASA and ESA archives.

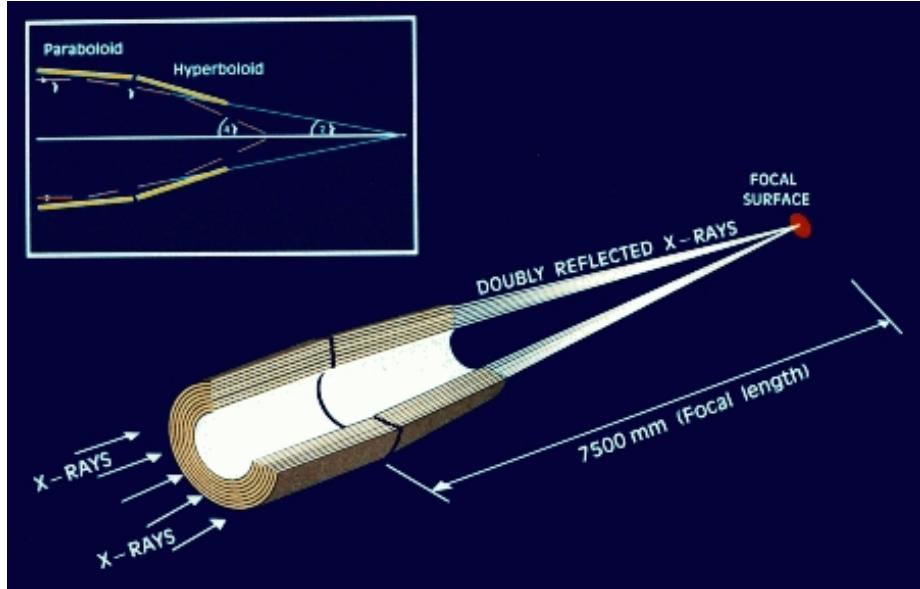


Figure 2.3: Grazing incidence reflection.³

2.1.3 Telescope Lens Properties

Since X-rays have such high energy, more common optical lenses cannot be used. The X-ray photons tend to penetrate straight through optical lenses instead of reflecting off them. This is why X-rays are also often used in medical environments since the X-rays can go straight through a human body and get stopped in denser places such as bones. Therefore, to focus X-rays, a different kind of lens is needed compared to more common optical lenses.

X-rays can reflect off metal surfaces as long as they approach them from a low enough angle. This is called grazing incidence reflection. Instead of reflecting the photons using a parabolically shaped mirror, such as the optical Hubble Space Telescope, the photons are reflected using grazing incidence reflection. The lens consists out of circular mirror segments that reflect the X-rays to the focal point, where the sensor is located, see Figure 2.3.

The small reflection angle causes a long focal length of 7.5 meters. This necessitates the long length of the *XMM-Newton* satellite of 10m. To maximize the collection area, multiple shells of concentrically arranged mirrors are used. Specifically, 58 gold-plated mirrors, with the diameter of the largest mirror being 70cm. [1] In Figure 2.4 one of these gazing-incidence lenses of the *XMM-Newton* can be seen.

Three elements characterize the performance of an X-ray telescope. Firstly, the ability to focus photons, expressed in the point-spread function (PSF). Secondly, the ability of the mirrors to collect X-rays at different energies, expressed

³Source: ESA <https://www.cosmos.esa.int/web/xmm-newton/technical-details-mirrors>



Figure 2.4: One of the three gazing-incidence lenses of the XMM-Newton.⁴

in the effective area. Thirdly, the ability to reject X-rays that took an undesired route through the lens and hit the sensor (straylight). The third element is less relevant is less relevant for our research and will not be discussed further.

Point Spread Function (PSF)

The point spread function (PSF) describes the response of the lens to a point source. When the *XMM-Newton* observes a point source, the lens is not able to perfectly project this on the sensor, and it instead smears the point out to a *blob*. The size of this *blob* is called the PSF and is measured by the size of its full width at half maximum (FWHM). If the two sources are close together, the PSF of these sources can overlap. In that case we cannot determine if we are observing one or two sources. Therefore, the PSF determines the resolution of the telescope.

The PSF of a source is mainly determined by the off-axis angle and the energy of the X-ray entering the telescope. X-rays with a larger energy cause a bigger PSF. The off-axis angle is determined by the distance from the boresight (the focal point). Due to the properties of the grazing-incidence mirrors, the PSF stretches with a bigger off-axis angle. The PSF of the *XMM-Newton* EPIC-pn is demonstrated in Figure 2.5.

Vignetting

Similar to the shape of the PSF, the effective area of the mirrors is also a function of the off-axis angle. With an increasing off-axis angle (further away

⁴Source: ESA <https://www.cosmos.esa.int/web/xmm-newton/technical-details-mirrors>

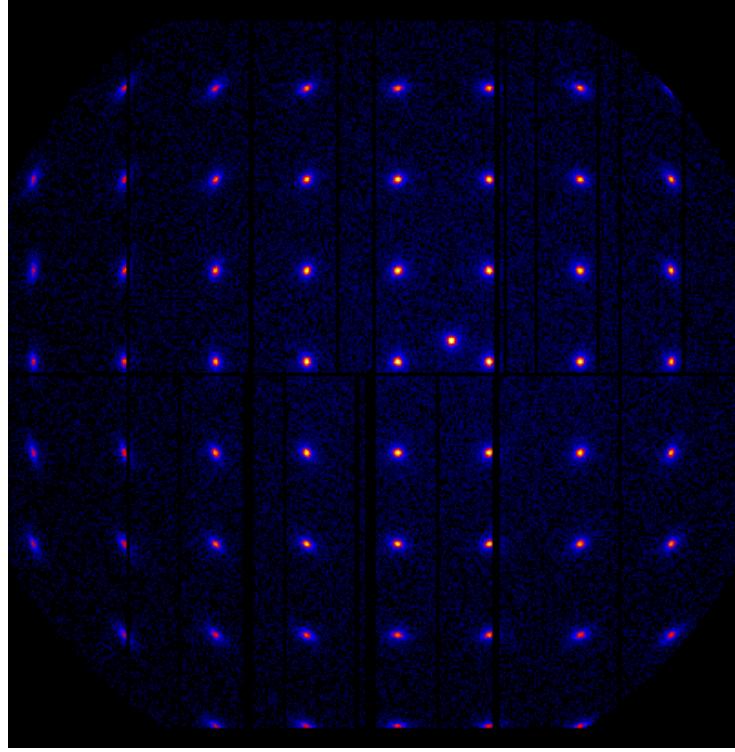


Figure 2.5: A grid of equally spaced point sources with the same brightness to demonstrate the *XMM-Newton* EPIC-pn PSF. One point source near the center is not part of the grid, but is located on the boresight. The black lines originate from the CCD-gaps. This image was generated using the simulator developed for this thesis.

from the boresight), fewer X-rays entering the telescope reach the sensor. This effect is called “*vignetting*”. This means that a source of the same brightness will be dimmer towards the edges of the image since fewer photons from this source reach the focal plane.

2.1.4 Telescope Sensor Properties

The sensor of XMM-Newton EPIC-pn (Figure 2.6) consists of 12 separate CCDs (charge-coupled devices), as can be seen in Figure 2.7, with each sensor being 64x200 pixels. The choice of separate CCDs was for redundancy reasons since once the telescope was launched, there was no possibility of repairs once something failed. In-between each sensor, there is a small *CCD gap*, which are also present in EPIC-pn observations as black lines. Since at a *CCD gap* no information can be recorded. Each pixel of the CCDs has a size of 150x150 microns, which corresponds to 4.1 arc seconds (≈ 0.001 degrees) and a position resolution

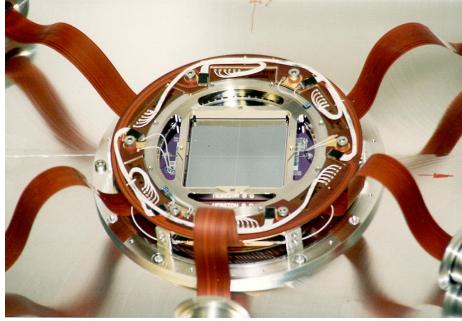


Figure 2.6: The *XMM-Newton* EPIC-pn sensor.⁵

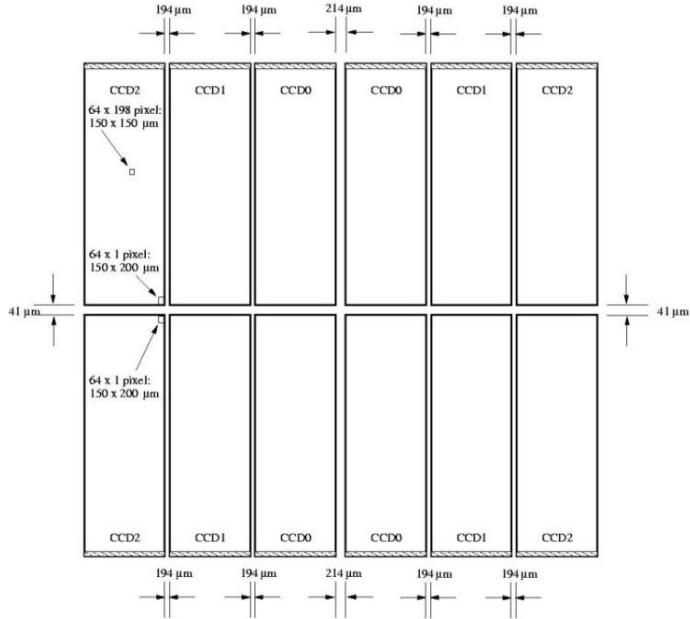


Figure 2.7: EPIC-pn CCD layout.⁶

of 120 microns; this results in an angular resolving resolution for a single X-ray photon of 3.3 arc seconds (≈ 0.0009 degrees). This means that the telescope can locate the position of an X-ray with an accuracy of 3.3 arc seconds (≈ 0.0009 degrees).

The boresight of the telescope is not focused on the center of the sensor, as can be seen in Figure 2.5. The reason for this is that at the center of the sensor, a *CCD gap* is present. Therefore, the sensor is slightly moved off-center such that the boresight hits the bottom center of CCD0 in the first quadrant.

⁵Source: ESA. <https://www.cosmos.esa.int/web/xmm-newton/technical-details-epic>

⁶Source: ESA. <https://www.cosmos.esa.int/web/xmm-newton/boundaries-pn>

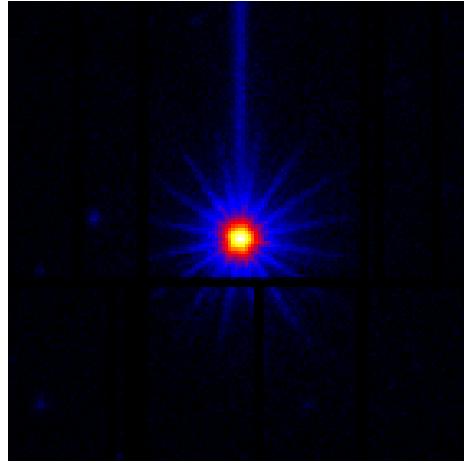


Figure 2.8: Observation of 2S0921-630, a bright X-ray source. The image center cropped around the boresight and log scaled to highlight the out-of-time events and telescope support beams.

Out of Time Events

The recording of an X-ray hitting the sensor, also called an event, works as follows: First, an X-ray hits a pixel of the CCD. The energy of this X-ray hitting the pixel generates an electric potential (charge). Based on the voltage of this electric potential, the energy of the X-ray can be determined. In order to do this, the electric potential has to be read out. However, the EPIC-pn sensor only has readout nodes at the edges (gray areas in Figure 2.7). So in order to get the electric potential to the edge, it is transferred to the neighbouring pixel towards the readout node. This way, the charge *jumps* to the readout, one pixel at a time. At the readout node the charge is measured and digitally stored. This process is very fast and usually does not interfere with the observations. However, when an X-ray hits a pixel while these charges are *jumping* towards the readout, an incorrect measurement occurs. The sensor will assume that the charge caused by the X-ray is actually a charge from another pixel that is *jumping* towards the readout. Therefore, it will record the wrong pixel location of the event. This is called an out-of-time event. This happens more often when looking at a very bright source since the sensor receives more X-rays. The incorrectly recorded locations of these X-rays are distinctly visible as a line towards the readout note, see Figure 2.8. These out-of-time events cause the line from the source to the edge of the sensor. In Figure 2.8 we can also observe smaller lines radially outwards from the source center, caused by the support-beams of the lens. Note that the image in the figure is scaled with the log function to highlight these phenomena. These phenomena are only significant with bright sources.

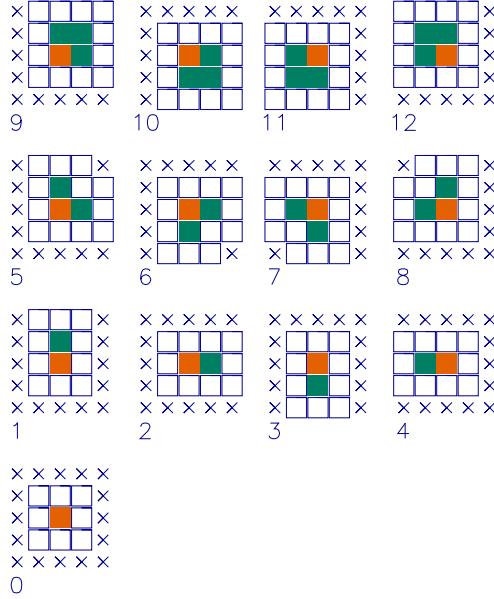


Figure 2.9: Patterns 0-12 of the *XMM-Newton* EPIC-pn sensor. Where the orange pixels have the highest charge and are the center of the event, the green pixels have a charge above the detection threshold, the white pixels have a charge below the detection threshold and the crossed pixels can any charge.⁷

Patterns

When a very energetic X-ray hits a pixel on one of the CCDs, it can cause such a big charge that the charge in surrounding pixels is affected. However, these very high-energy X-rays are mostly caused by solar flares [30] and do not originate from the source of interest. It is, therefore, useful to detect and filter out these events. A good X-ray sensor reading is compact, consisting of a high charge at the center and a low charge in the surrounding pixels. If the surrounding pixels also have a high charge, it might be from a solar flare X-ray. We want to filter this out. The distribution of these surrounding charges can be visualized in patterns, see Figure 2.9. To filter out these spurious high-energy X-rays, generally, only single and double patterns (≤ 4) are considered.

2.1.5 Background Noise

All *XMM-Newton* observations have background noise. The relation between the signal of the source of interest and the background noise (signal-to-noise ratio) determines the quality of the observation. Therefore any techniques to

⁷ Adapted from <https://xmm-tools.cosmos.esa.int/external/sas/current/doc/emevents/node4.html>

reduce the background noise, such as de-noising, are highly valuable. This background noise in *XMM-Newton* observations originates from:

- X-ray photons:
 - Astrophysical background, at lower energies mainly caused by thermal emission ($E < 1$ keV) and at higher energies mainly from unresolved cosmological sources. These sources are so far away or so faint that the number of photons that will hit the telescope will not be enough to resolve it.
 - Solar flares.
 - Reflections from outside the field-of-view .
 - Out-of-time events.
- Particles:
 - Soft proton flares. These low-energy protons (less than 100 keV) are reflected onto the sensor by the telescope mirror. When these protons hit the sensor, they cause a charge similar to X-ray photons.
 - Internal particles. These are particles that penetrate the sensor from any direction.
 - X-rays created by particles. A particle can hit the physical satellite material and create an X-ray, which can then hit the sensor.
- Electronic sensor noise:
 - Bright malfunctioning pixels and columns.
 - Readout noise caused by thermal energy.

There are many methods to filter background events, including filtering the earlier described patterns or excluding events during a known solar flare.

2.1.6 Image Generation

During an observation, every event is stored in an eventlist. For every event, the pixel location, energy, patterns, and recording time are recorded. To generate an X-ray image, we select the events we are interested in and count the corresponding pixel values. These pixel values are then saved in a fits image. The fits format [31] is a lossless file format with an uncapped data range. This is important since thousands of events can be recorded at one pixel. Common image file formats are usually only able to store 255 discrete values.



Figure 2.10: Direct demodulation example on supergiant galaxy M87. From *left* to *right*: *XMM-Newton* EPIC-pn image (pixel size = 2.5 arcsec), restored image using DD (pixel size = 0.5 arcsec) and *Chandra* ACIS-I image smoothed using a Gaussian filter (pixel size = 1.2 arcsec). Adapted from [15].

2.2 Super-Resolution and De-Noising Methods

At the time of writing, no published research has been done on applying machine learning-based super-resolution or de-noising on *XMM-Newton* observations. A few non-machine learning-based methods will be discussed in this section.

2.2.1 Super-Resolution Methods

Traditionally, SR uses interpolation methods such as bilinear and nearest neighbour interpolation for upscaling. However, these methods often introduce side effects such as noise amplification and blurring. A method that suffers less from these problems is direct demodulation (DD) [32]. This method is mathematically complex; therefore, we will not go into too much detail on its inner workings. Some key steps concepts include:

- Estimating the background noise around the source of interest.
- Converting the image to a correlating matrix.
- Iterative solve the correlation matrix with a formula based on the PSF of the telescope constraint on the background.

The requirement of estimating the background noise around the source of interest means that this method has to be manually tuned for every source of interest. It can, therefore, not automatically be applied to observations.

Using the DD method [15] demonstrates that the spatial resolution of *XMM-Newton* EPIC-pn images can be improved by a factor of 5 while adhering to the requirements for spectral studies. An example of the DD method applied on *XMM-Newton* EPIC-pn data is shown in figure 2.10.

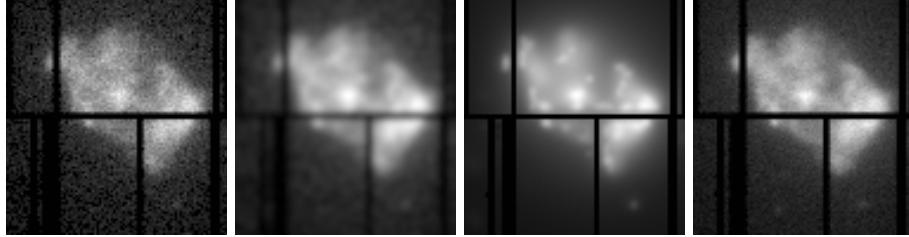


Figure 2.11: Gaussian and wavelet transformed example on supernova remnant SNR 3c397 observed with *XMM-Newton* EPIC-pn. Scaled with the square root function for visualization. From *left* to *right*: Original image at 20ks exposure, Gaussian blurred with a $r = 3$ and $\sigma = 1.5$, wavelet transformed and a reference image at 119ks exposure.

2.2.2 De-Noising Methods

Non-machine learning-based de-noising methods include Gaussian filtering and wavelet-based filtering. Gaussian filtering works by convolving the image with a Gaussian kernel that has a certain radius (r) and standard-deviation (σ).

Wavelet-based filtering [33] works by decomposing an X-ray image with an un-decimated discrete wavelet transform. This produces a series of images containing the signal present at different scales in the image. A model for Poisson noise is used to isolate which wavelet coefficients (i.e. pixels at each scale) exceed a given significance compared to a flat signal distribution. Finally, the output image is reconstructed iteratively from only the significant wavelet coefficient. The use of wavelet-based de-noising methods has been shown to optimize the detection of AGNs, galaxy clusters and other features in X-ray images of different X-ray telescopes [34][35][36].

Figure 2.11 gives an example of the Gaussian filtering and wavelet transform de-noising methods on a *XMM-Newton* EPIC-pn observation.

Chapter 3

Data

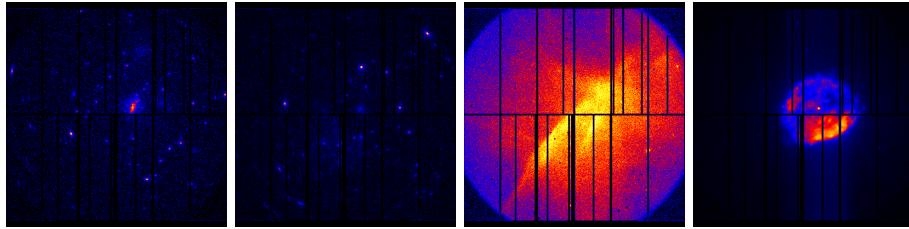


Figure 3.1: Example images from the real *XMM-Newton* dataset scaled with the square root function to visually enhance pixels with lower counts. From *left* to *right*: NGC 4666 with 53ks exposure, M101 with 71ks exposure, Vela SNR with 59ks exposure and 1E 1613-5055 with 56ks exposure.

To train and validate our models we created a dataset consisting of real XMM observations (section 3.1) and a separate dataset of simulated XMM observations (section 3.2).¹

3.1 Real XMM-Newton Dataset

XMM-Newton observations are in the form of *eventlists* that records the time photons of a certain energy hits a specific CCD pixel. We need to transform these eventlists into images, so in order to limit the scope of this research we focus on the PN-detector in the (extended) full-frame mode and the photon energy in the band between 0.5 and 2.0 *KeV*. Many interesting observations of extended sources where structure such as shock waves are made visible with this detector, filter, and energy band.

We use the entire *XMM-Newton* archive, filtering out observations with less than 20 ks exposure times, bad time intervals and events. We split the event

¹Source code available at: <https://github.com/SamSweere/xmm-epicpn-simulator>

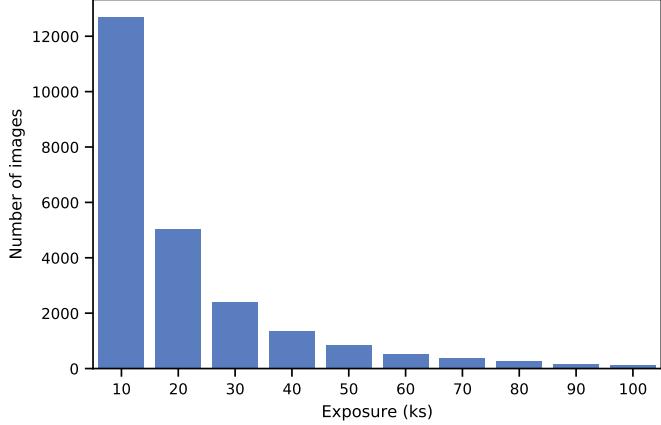


Figure 3.2: The distribution of image exposure times from the real dataset.

lists in 10 ks intervals for each observation, i.e. for a 40 ks observation, we generate 4x10 ks images, 2x20 ks images, 1x30 ks and 1x40 ks image. The images with multiple exposure times enables us to train super-resolution and de-noising models using the same exposure-time for different observations. It also enables us to train a de-noising model with pairs of low and high exposure images. The exact implementation details of generating the real XMM-Newton dataset are described in appendix A.1.

The final dataset contains 5554 unique images (see e.g. Figure 3.1 and their exposure time distribution after splitting ranges from 10 to 100ks (Figure 3.2).

3.2 Simulated XMM-Newton Dataset

The real XMM-Newton dataset cannot be used to train a super-resolution model. For super-resolution, we need to create a dataset consisting of low input resolution images and their high resolution counterparts as our targets. This is often achieved digitally, through down-sampling high resolution images [19][21][37][38] or optically, through aligning images taken with different zoom scales on the camera lens [37][38][17].

These methods are not applicable to our research as will be explained in this section. In order to create training pairs we therefore decided to use simulation software, where we can artificially increase the resolution, both the angular resolution and the sensor resolution. The motivation for building a simulation setup was due to easier solutions not being sufficient. In this section we go through a few of these options such as down-sampling and zooming and explain why they would not work for our research.

Down-sampling

Down-sampling the real XMM images would not be useful, since the image properties do not stay the same when downsampling. This is often not a problem with other types of images since these are somewhat scale invariant, i.e. the object in the picture looks the same when taken at different resolutions. This is not the case for XMM observations, a few reasons for this are:

- CCD layout and bad pixels are always at the same position, down-sampling would change the size of these.
- Point sources such as AGNs have a specific PSF. Downsampling an image would make this PSF smaller.
- This PSF changes varying based on the location in the picture (as described in section 2.1.3).
- The background noise variation would decrease when downsampling, making the background noise different to the real observations.

A model trained on such a dataset would most likely learn the transformation between the down-sampled and the real XMM images. However, the transformation from a real XMM image to the super-resolution image is different and such a model would therefore not be applicable.

In astronomy, we require higher resolution both spatially and in terms of the PSF. The XMM PSF is spatially variant, and should be smaller for a higher resolution observation. Which is as explained not possible with down-sampling.

Zooming

Another way to generate low and high resolution image pairs is by zooming in using the lens of a camera [17]. This is not an option however in the case of XMM, since the lens (a glancing reflector) cannot zoom. Another option would be to use the data of another X-ray telescope that has a higher spatial resolution such as the Chandra X-ray Observatory for the high resolution images similar to [28]. One could then use the XMM image of a specific source as the low-resolution input image and the Chandra observation of the same source as the high-resolution target image. This option was considered, however there are a few key problems when using this approach. The first problem is that there are a limited amount of interesting observations of the same source from both XMM and Chandra, most likely not enough to properly train a model on. The other problem is that they are different telescopes with different properties. For example Chandra has a higher spatial resolution but is less sensitive [39], certain structures that are present in the XMM observations will thus not be visible in the Chandra images. This would limit the potential of the super-resolution images of XMM.

Limited Number of Extended-Sources

Another potential problem is that we are interested in sources that contain a lot of structure, such as super-nova remnants with shockwaves. However, there is only a limited amount of observations of such objects. This will limit the amount of available training, validation and verification data. Thus restricting the training of our model on generating SR and DN images containing these sources.

3.2.1 Our Approach

For this research, we require a dataset where the high and low resolution pairs have the same image properties. As mentioned before, this is not feasible with real data so a simulation approach was chosen, where we can artificially increase the resolution (both the angular resolution and the sensor resolution) whilst maintaining the relevant observational properties.

For this research we decided to use simulation software to create a simulated dataset. We build this software based on the SIXTE X-Ray simulation software package [40]. To make this work for our use-case, the *XMM-Newton* EPIC-pn sensor with a high variety of structured sources, we had to create a lot of custom configuration files as well as creating the input sources. The rest of this chapter will be dedicated to how the simulation was developed to create the simulated dataset.

3.3 Simulation Software Setup

One existing *XMM-Newton* simulator is SCISIM [41], however it is limited by the number of simulated images it can create. Therefore we chose to use the SIXTE software package [40]. This is a X-ray simulation software package provided by ECAP/Remeis observatory². The default *XMM-Newton* configuration files, were not suitable for this work and therefore we create custom configuration files to simulate the *XMM-Newton* epic-pn sensor in the 0.5 to 2 KeV energy band. This configuration provides a realistic replica of the telescope and sensor properties.

We create two sets of simulations. The first simulation is made to mimic real XMM observations, and the second simulation containing twice the resolution. Here, the lens is twice as accurate, which means that the PSF of a point source is halved. In order to record this improved lens resolution we also need twice as many pixels on the sensor. For a detailed description of the simulation configuration see appendix A.2.

²<https://github.com/thdauser/sixte>

3.3.1 Simulated Telescope Lens Configuration

One of the key elements we have to simulate is how the light from a source reaches the sensor, i.e. the lens of the telescope. The construction of the *XMM-Newton* lens is different than optical cameras, as explained in section 2.1.3. These unique properties need to be simulated as well in order to get simulated images that are as close as possible to real XMM observations.

PSF

The simulator needs to simulate the behaviour of the real telescope, as well as a telescope with twice the spatial resolution. With twice the spatial resolution we simulate a PSF that is twice as small. The implementation details are described in appendix A.2.2.

Vignetting

The vignetting properties of the *XMM-Newton* are well documented, in the simulator we therefore use a recent calibration file ³. This works for both the normal and higher resolution simulations because it is not influenced by the PSF or sensor resolution.

3.3.2 Simulated EPIC-pn Sensor Configuration

The *XMM-Newton* sensor has specific properties that we need to simulate to get a realistic image. However, not every sensor property significantly impacts the final image, also taking into account the kind of sources we are simulating, such as the sensor behavior when looking at very bright sources. We, therefore, decided to focus on implementing the properties that have the most influence on our observations.

CCD Layout

The EPIC-pn sensor consists out of 12 separate CCDs, as described in section 2.1.4. Initially, we simulated all these sensors separately based on the physical properties, including out-of-time events. However, this significantly increased the computation time of the simulator since in *SIXTE* the whole telescope has to be separately simulated for each sensor. Thus, to get one observation, 12 separate simulations have to be run. The benefits of simulating every sensor separately are that specific properties such as out-of-time events can be simulated as well. However, the sources we are interested in are not extremely bright. These sources, due to their lower brightness, cause less or no out-of time-events. Furthermore, we argued that the impact of not having these properties would be minimal on the final image. Therefore, we decided to simulate all 12 sensors as one big sensor, and thus not simulating the physical properties that cause out-of-time events. Since we now do not have any chip gaps in-between the

³XRT3_XAREAEC_0012.CCF

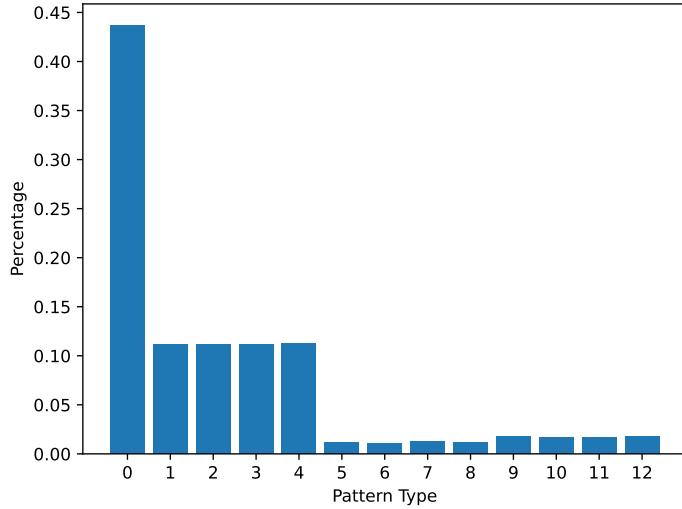


Figure 3.3: The pattern distribution of an unfiltered combined eventlist of the test grid using all CCDs.

CCDs, we multiply the final image with the detector mask. The detector mask filters out all the areas in the image where no recording of events is possible, including the chip gaps, known faulty pixels, and areas outside the field of view. This sped up the simulation approximately 12 times.

Patterns

Very energetic X-rays readings can affect the pixels around it. However, very high-energy X-rays are likely caused by sources other than the one we are observing, such as solar flares. To filter out these readings, we look at the pattern of measurements in surrounding pixels (for more details, see section 2.1.4). In Figure 3.3 we can see the patterns that are observed with the simulator when observing an artificial test grid of point sources(Figure 2.5). Note that this observation is relatively bright and therefore has relatively high presence of higher pattern types. For our final images we only keep single and double patterns, i.e. pattern ≤ 4 .

3.4 Simulation Input

XMM observations can be categorised into three parts: the extended source, point sources, and background noise. We structure our simulation input (simput) into the same three categories. The final format of these is SIMPUT (SIMulation inPUT) [42].

3.4.1 Extended Source Component

For the simulation input of the extended source, we need something realistic, and of the type we are interested in, such as supernova remnants and galaxy clusters. Therefore, we chose to use results from the IllustrisTNG suite, which contains these types of sources.

The IllustrisTNG suite consists of a series of large, state-of-the-art cosmological magnetohydrodynamical simulations of galaxy formation. They simulate a mock Universe starting from soon after the Big Bang until the present day, taking into account a large range of physical processes that drive galaxy formation. The three simulations we are using are IllustrisTNG 50-1 [43][44], IllustrisTNG 100-1 and IllustrisTNG 300-1 [45][46][47][48][49]⁴. These are simulated at different scales (cubic volumes of ~ 50 , 100 , and 300 Mpc side lengths) and mass resolutions that enable the study of different types of sources - supernova remnants at the smaller scale and galaxy clusters at the larger scale. The simulations include full baryonic physics. In each simulation, we select the top 400 subhalos (clusters of mass) based on the M_{gas} . We then project the subhalo from the x,y, and z axes on two different scales. We project at two different scales for a close-up of the source and a projection that is four times further away (tng50-1: 100 and 400, tng100-1: 400 and 1600, tng300-1: 1000 and 4000 physical kpc) in order to capture different spatial information from the same source. From these projections, we calculate the X-ray photon intensity in the 0.5-2.0 kev energy range at redshift 0.01. Finally, we save these projections to a simput (simulation input) image that will be used as an input to our *XMM-Newton* simulation. The resolution of this image is 2048x2048, which is bigger than the sensor resolution of the simulated *XMM-Newton*, both at 1x and 2x scales. This way, the simput image resolution is not a bottleneck for our simulation. Some subhalos emit almost no X-rays, look like point sources at our scale, have too low resolutions, or are incomplete. These will not be interesting for our purpose and are therefore manually removed from our dataset. This results in:

- TNG50-1: 1632 images
- TNG100-1: 2165 images
- TNG300-1: 2374 images

⁴<https://www.tng-project.org/>

Flux Determination

The flux (brightness) of the extended sources are modified to reflect real extended sources that XMM observed. We set the central part of the source (a box at 5% of the image width/height at the center of the image) to have a flux value randomly sampled uniformly to be between 5 and 50 times the standard deviation of background noise (σ_B) at the boresight. The sources are modelled as a power law spectrum using *xspec* [50] with $n_H = 0.04$, $PhoIndex = 2.0$ and $norm = 0.01$.

Spatial Augmentation

Additionally, to increase our training sample size, we artificially augment the data⁵ by apply a random zoom of scale [1, 2]. Here we made sure that at the highest zoom scale, the resolution of the final simput image is still larger than the maximum XMM simulated resolution. Such that this does not form a bottleneck in our simulation pipe-line. We also do a random x, y perturbation offset. Most real XMM observations have the source of interest at or around the boresight. Therefore, we also keep the center of our source close to the boresight by using a standard distribution with a standard deviation of 5% of the image height/width. This offset is limited to the center $\pm 1/3$ of the image size, such that the edge of the simput image is never too close to the edge of the XMM simulation image.

Final Extended Source Simputs

For every unique projection of a subhalo we generate five simulation inputs (simputs). These five inputs are augmented with different zoom ranges, fluxes and spatial offsets as described above. After the various augmentations are applied we are left with:

- TNG50-1: 8160 simputs
- TNG100-1: 10825 simputs
- TNG300-1: 11870 simputs

Thus in total 30855 augmented simputs that are used as the extended source component of the simulated XMM observations. Figure A.1 shows three examples.

3.4.2 AGN Component

The point source component of the observations are mostly AGNs. AGNs (active galactic nucleus) are compact regions at the center of galaxies that have a very high luminosity. Given their compactness they are usually only visible as point

⁵Note that these are not the final simulated images but the extended source components that will be used to input our XMM simulator.

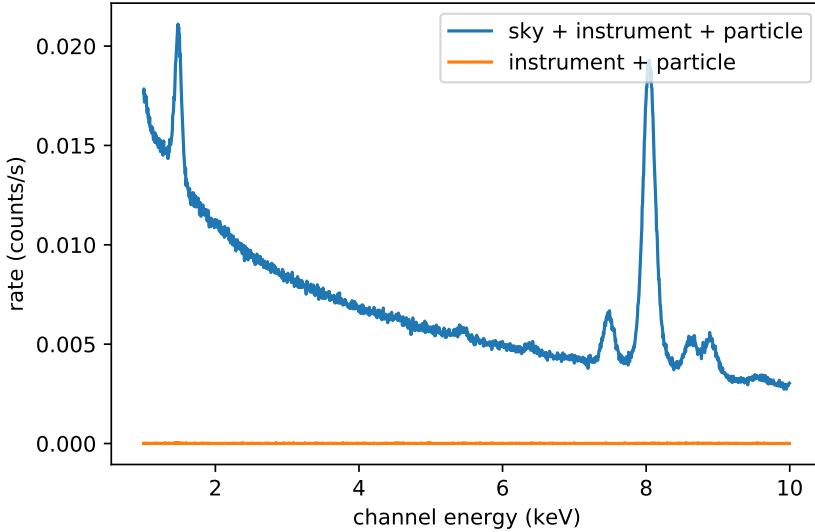


Figure 3.4: Background noise in counts per second for every energy channel.

sources. Based on measurements of AGN number counts N and their relative fluxes S [51], we compute the expected number of AGNs in each simulated observation corrected for the XMM FOV. [51] do not publish their uncertainties, therefore we include an additional Poisson uncertainty $\pm\sqrt{N_{\text{stars}}}$.

AGN's are present in the whole sky, however, the galactic plane (our own galaxy) partially absorb the X-rays from far away AGNs. The distribution of the AGNs is therefore also dependent on the galactic latitude. To simulate this we set the absorption rate of the galactic plane (0 degrees) to $1 \cdot 10^2$ and no absorption when looking out of the galactic plane (90 degrees). Real observations that use the EPIC-pn full-frame observation mode, tend to be pointed away from the galactic plane to avoid the bright sources that are common there. We therefore, sample the absorption linearly in logarithmic space to favor out of plane observations.

We simulate absorption by shifting the $\log(N)/\log(S)$ distribution and dividing S by the absorption rate. We make a flux cut at $1.5849 \cdot 10^{-15} \text{ erg/cm}^2/\text{s}$, such that any source fainter than this will not be simulated, and we require at least one observable AGN in each observation.

3.4.3 Background

All *XMM-Newton* observations contain noise. This noise consists out of cosmological background noise which mainly originates from thermal emissions and unresolved cosmological sources. Another contributor to the noise is not gener-

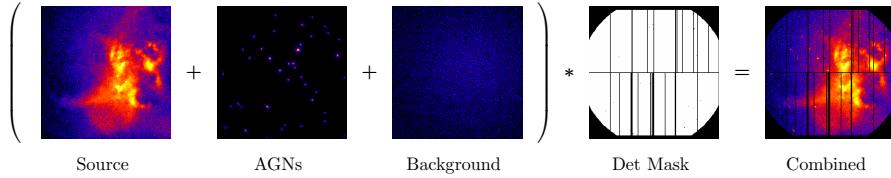


Figure 3.5: Different components that make up a simulated XMM observation. The simulated extended source, AGNs and background are added together and then multiplied by the detector mask to create a simulated XMM observation. The images are logarithmic scaled for visualisation.

ated by X-rays but for example caused by protons and instrument noise. This is explained in more detail in section 2.1.5. For our simulator we also have to simulate this noise to create realistic observations.

To simulate this background noise we use the prepared blank sky eventlist⁶⁷ `pn_t_ff.g` [30] to extract the background noise at all energies. The distribution is shown in Figure 3.4. Since the instrument + particle background is minor with respect to the sky + instrument + particle background we do not simulate this. The peaks at certain energy levels correspond with the spectral emission of specific elements present in background sources.

3.4.4 Simulated XMM Training Data

The 3 observational components (source, AGN and background) are simulated separately and combined using addition to create the final XMM observation, as can be seen in Figure 3.5.⁸ Simulating these separately enables us to create many combinations of observations. For example, if we look at uniquely simulated sources at 100ks, we have 30855 simulations of extended sources, 25000 simulations of point source backgrounds, and 25000 simulations of background noise. This amounts to $30855 \cdot 25000 \cdot 25000 \approx 2 \cdot 10^{13}$ possible unique simulated XMM observations. We generate pairs of noisy XMM observations to use as inputs to our network, and high-resolution, noiseless images as our network targets. Although many of these will be visually similar, the different combinations help reduce over fitting of the model. This compensates for the lack of traditional image augmentations such as spatial and colour transforms. Spatial augmentations are not usable since the PSF, vignetting and chip-gap layout are spatially dependent. Colour augmentations would change pixel values to something that does not resemble realistic observations. The final combined simulated XMM observation has the same image proprieties (dimensions, type)

⁶Observations where all the discrete sources, such as extended sources and AGNs, have been removed.

⁷https://xmm-tools.cosmos.esa.int/external/xmm_calibration/background/bs_repository/blanksky_all.html

⁸It took approximately 6000 CPU hours to complete these simulations on a compute cluster.

as the real XMM observations at the 1x scale. At the 2x scale the spatial resolution is double and the PSF size is halved. Figure 5.1 shows three examples.

3.4.5 Train, Validation and Test split

The final simulated images are split into train, validation and test sets where only the training dataset is used to update the weights of the network, the validation data is used to monitor the performance of the network and the test data is reserved for final evaluation of network. Both validation and test data are not seen by the network during training.

For the simulated dataset, the splits are made in a way that all the spatial augmentations done during the simulations are always in the same set. Note that a specific source can appear multiple times across the sets but with different projections and distances. The choice to not split based on the sub-halos themselves was made because rare source structures could then be over-represented in one of the splits. Since different projections and distances of the same source look very different, the risk of the model memorising these sources should be minimal. The train, val, test split distribution chosen for the simulated dataset is 80, 10, 10.

The train, val, test split distribution chosen for real XMM observations is 70, 15, 15. Since the real dataset has a smaller number of images we need a larger percentage of the images to validate and test the results in comparison to the simulated dataset.

Chapter 4

Experimental Methodology

In this chapter we explain the methodology of developing our final models. Consisting of the data prepossessing (section 4.1), the architectures of the super-resolution and de-noising models (section 4.2), data and model hyper-parameter tuning (section 4.3) and finally the process of training the final models (section 4.4).¹

4.1 Data Prepossessing

To help accelerate the optimization of the model and more efficient convergence, the data need to be prepossessed. We transform the data from *counts* to *counts/s* by dividing the image by the exposure time. This enables use training data with different exposure times whilst maintaining the input pixel range to our model. Bright sources can have large pixel values that are orders of magnitudes higher than other pixels in the observation. This big difference can make training a deep learning model very unstable and therefore we clip pixel values to 200 times the mean background rate $\mu_B = 1.1168 \cdot 10^{-5} \text{counts/s}$ for the de-noising dataset and $50\mu_B$ for the (2x) super resolution dataset. The clipping value for 2x is four times smaller because the pixel density is four times larger, meaning that the pixel counts on one pixel on the 1x resolution scale will be distributed over four pixels in the 4x resolution scale. This can lead to the loss of detail in bright sources, such as AGNs. However, the majority of the extended features have X-ray counts below 200 times the mean background. Finally, the image is then normalized to $[0, 1]$, where 0 is no counts and 1 is the 200 times μ_B .

4.1.1 Data Scaling

The number of photon counts of different *XMM-Newton* observations can vary by orders of magnitudes. A bright AGN might be responsible for 1000 counts

¹Source code available at: <https://github.com/SamSweere/xmm-superres-denoise>

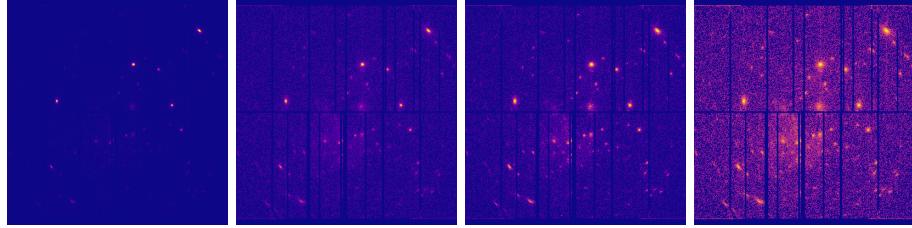


Figure 4.1: Examples of the different scaling applied to galaxy M101 (obs id: 0824450501). From *left* to *right*: linear, square root, asinh and log.

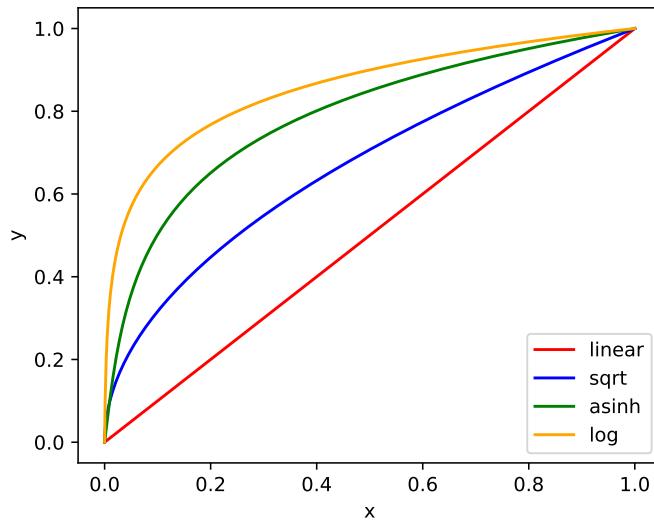


Figure 4.2: Visualisation of the data scaling functions. With input x and scaled output y .

whereas the edge of a supernova remnant only 10 counts. These smaller features would not be visible to the human eye when visualizing these images without data scaling (see Figure 4.1).

It can also affect the training of the model. For example, an L1 loss, would put more weight on features with higher pixel values and bias the results. Our main focus is to enhance the visual clarity of smaller features, and for this reason we explore several different data scaling functions. We compare linear, square root, log and asinh scaling functions (see Figure 4.2).

Many interesting structures have pixel counts a few times bigger than the standard deviation of the background (σ_b), while bright parts of the image such as the center of agns can have pixel counts in the hundreds of σ_b . Since we are interested in the structures, we stretch the pixel values by taking the image's square root (sqrt). This causes smaller values to increase relative to bigger values. Note that the pixel value range [0, 1] before and after the sqrt scaling

remains unchanged. Similarly this is true for the linear scaling. However the remaining scaling functions will alter the range of pixels of our input data and therefore we modify them to maintain the pixel range.

Log is a commonly used scaling function in astronomy since the human eye responds to light in a logarithmic manner. We define our log scaling as:

$$y = \frac{\log(ax + 1)}{\log(a)}, \quad (4.1)$$

where we empirically set $a = 1000$.

Lastly, the asinh scaling offers us more flexibility than the log function. Our asinh scaling function is defined as:

$$y = \frac{\operatorname{arcsinh}\left(\frac{x}{a}\right)}{\operatorname{arcsinh}\left(\frac{1}{a}\right)}, \quad (4.2)$$

where we fix $a = 0.05$ such that it visually highlights features of a specific brightness. As defined, this function lies between the *log* and *sqrt* scaling functions (see Figure 4.2).

4.2 Super-Resolution and De-Noising Model

As stated previously, super-resolution (SR) and de-noising (DN) are ill-posed problems. Given a low-resolution and noisy image, there are multiple possible SR and DN images since the image does not contain enough information to generate the definite details for the SR and DN image. For example, it is impossible to predict the background noise for a SR and DN image based on the low-resolution and noisy image since the background noise is inherently random.

Originally, for our super-resolution problem, we took an approach based on a GAN architecture [26]. GANs use a generator network to generate realistic images and a discriminator network to ensure that the generated images are visually indistinguishable from the high-resolution target images. We chose the ESR-GAN model [19] for its proven success as a super-resolution model. It consists of a stacked Residual-in-Residual Dense Block (RRDB) (see subsection 4.2.1) generator and a deep CNN discriminator. However, like all GAN based algorithms, ESR-GAN suffers from hallucinations of non-existent features in the model output. This can have catastrophic consequences in astronomy. For more robust reconstructions, and at the expense of generating images that are less visually similar to the target, we choose to omit the adversarial component. Our main model is therefore based on the RRDB generator used in ESR-GAN. This architecture generates more reliable outputs, however is only able to generate sharp reconstructions in areas of high confidence. Aspects of low confidence such as the background noise will be result in blurry reconstructions and the output image is unlikely to visually resemble the target.

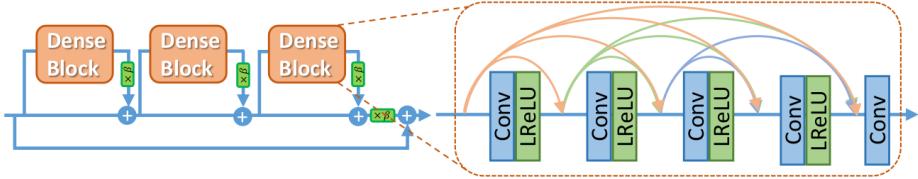


Figure 4.3: RRDB Basic Block, adapted from [19].

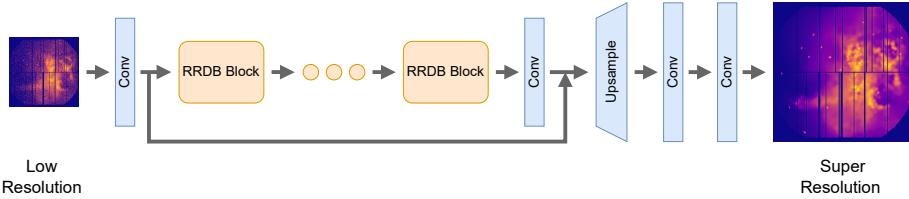


Figure 4.4: RRDB Super-Resolution Model Architecture. The network takes a low resolution image and undergoes a convolutional layer followed by a series of RRDB blocks, another convolutional layer with skip connections, an upsampling layer, and finally two more convolutional layers to return a higher resolution mapping.

4.2.1 Model Architectures

The main feature of our architectures is the use of the RRDB block (Figure 4.3). This block is inspired by the DenseNet architecture [52] and connects all layers within the residual block with each other. It consists out of three Dense Blocks, within which contain 4 consecutive convolution layers followed by Leaky ReLU activations and an additional convolutional layer. The concatenated output of every previous layer is fed into the next convolution layer. The number of convolution filters in every consecutive convolution layer increases linearly:

$$n_{i+1} = n_i + n_0 \quad (4.3)$$

Where n_{i+1} is the number of convolution filters in the next layer, n_i is the number convolution filters in the current layer and n_0 is the number of convolution filters in the first layer.

For our task of super-resolution, we base our architecture on the original ESR-GAN generator (Figure 4.4), however we replace the nearest-neighbour interpolation upsampling layer with pixel shuffle upsampling [53]. Pixel shuffle has more connections and does not interpolate the upsampled image. This should improve quantitative details on smaller scales.

For the de-noising model, we use the architecture however we remove the upsampling layer and the last convolutional layer (Figure 4.5). Additionally more skip connections are introduced as inspired by [54]. This helps to learn smaller features in the image and improves training speed since the model does

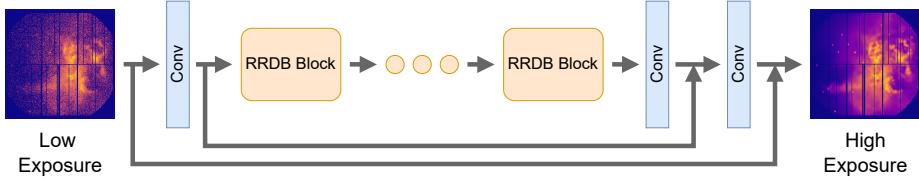


Figure 4.5: RRDB De-Noise Model Architecture. The network takes a noisy low exposure image and undergoes a convolutional layer followed by a series of RRDB blocks, another 2 convolutional layer, with skip connections to output a higher signal-to-noise mapping.

not have to process all the small features. Instead, it only learns the features to suppress to create the de-noised output image.

Weight Initialisation

Since our output images have values between 0.0 and 1.0, we need to ensure that the first pass through the model results in values in this interval. If this does not happen, when for example the values all are negative, everything will be clamped to 0.0. This will result in no usable gradients for backpropagation, i.e. the model will not train. Therefore we skew the initial weights in the last convolution layer to be slightly more positive. Normally the weights are initialized using a random normal distribution where the standard deviation is based on the size of the convolution layer:

$$std = \frac{1}{\sqrt{layersize}} \quad (4.4)$$

The weights are uniformly distributed from $[-std, std]$. In order to prevent the initial forward passes from being outside the image range, we initialize the weights in the last convolution layer with an increased higher bound of the normal distribution by $0.01 \cdot std$. The weights are thus uniformly distributed from $[-std, std + 0.01 \cdot std]$.

4.2.2 Loss Functions

The loss function determines how good a prediction of the model is with respect to the reference image. In preliminary testing, we observe a substantial difference in the visual appearance of the generated images and the target images. However, we require a more quantitative measurement of the reconstruction than a simple visual comparison. Different loss functions, optimise the model for different attributes of the output. We consider L1, Poisson, Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) [55], and Multi-scale Structural Similarity Index (MS-SSIM) [56] loss functions.

The L1 loss minimizes the mean absolute difference between pixel values of the generated and target images. This is the simplest loss function. We do not

include the mean square error loss (L2) as it is sensitive to outliers and extreme values [57] which can lead to bad performance on e.g. observations of AGNs. We include the Poisson loss, which measures the likelihood of the generated pixel values assuming that the target comes from a Poisson distribution conditioned on the input. It's relevant here because our data is count data and follows a Poisson distribution. The PSNR is a measure of the ratio between the maximum signal and the distorting noise. It is one of the basic metrics in de-noising models. A higher PSNR value equates to better de-noising. Lastly SSIM and MS-SSIM are perceptual metrics that incorporate the idea that spatially close pixels have strong inter-dependencies. These losses therefore measure the similarity of structure in images on a single scale and a combination of different scales respectively.

We additionally test combinations of these loss functions, however, since the values of different loss functions can differ by orders of magnitude, we normalize the loss based on the trial runs of the model trained with Poisson loss and the various data scaling functions. This ensures the losses are approximately on the same scale. For the exact implementation details see section A.4.

4.2.3 Image Quality Metrics

To evaluate our models, we tracked a range of image quality metrics, each corresponding to a different visual qualities. Their implementation details are described in section A.3.

- L1: Mean absolute difference between pixel values of the generated and target image.
- PSNR: Peak-Signal-to-Noise-Ratio, is a measure of the ratio between the maximum signal and the distorting noise. This metric is often used in astronomy to determine the quality of an observation.
- Poisson: Likelihood of the pixel values of the generated assuming the target comes from a Poisson distribution conditioned on the input.
- SSIM [55]: Structural Similarity Index, instead of having the quality be based on the comparison of the individual pixel values between the generated and target image, this metric incorporates the idea that pixels that are spatially close have strong inter-dependencies.
- MS_SSIM [56]: Multi Scale Structural Similarity Index, a combination of SSIM losses calculated at different scales. This metric therefore gives a quality estimation of structure at smaller and larger scale.
- FSIM [58]: In a similar fashion to SSIM and MS-SSIM, FSIM is a metric that aims to mimic human vision. The human visual system perceives images through salient low-level features, and FSIM uses 2 kinds of these features to determine image quality - the phase congruency (PC) and the

gradient magnitude (GM). Rather than the areas with sharp changes in contrast, PC highlights features as areas where the order in the phase component of the Fourier transform is high. Thus PC is an illumination and contrast invariant measure of feature significance. However since contrast is also an important aspect of human vision, FSIM also incorporates gradient magnitude to encode contrast information.

- HaarPSI [59]: HaarPSI uses coefficients obtained from a discrete wavelet transform to construct local similarity maps between two images. The Haar wavelet is used, being the simplest and most efficient to compute. Next a non-linearity is applied in the form of a logistic function to highlight the relative importance of those areas.

We measured these metrics for every data scaling (linear, sqrt, asinh, and log), in order to have a qualitative measure of the scaled images. This means that we scale the generated and target image with the data scaling function and then calculate the metric. Since the data scaling changes the distribution of the pixel values the image quality metrics also have different values.

4.3 Hyper-Parameter Tuning

Hyper-parameters influence the training of a model and its performance. To tune for the optimal configuration we perform a parameters search where we trained many models with different hyper-parameters to gain insight into the influence of each hyper-parameter on the model performance. There are two categories of hyper-parameters: the model hyper-parameters (subsection 4.3.1) and the data hyper-parameters (subsection 4.3.2). The model hyper-parameters are tuned first and fixed before the data hyper-parameters are tuned. We focus our hyper-parameter tuning on SR and DN models.

We use a grid-search approach to hyper parameter tuning, which can be computationally expensive and therefore we only train on a 25% subset of the simulated dataset where the inputs are further cropped to 128x128 pixels around the boresight. We train the models for 50 epochs on this reduced dataset. Although this is slightly different from the final model training, we argued that it gives enough insight into the model performance to make informed choices on the hyper-parameters used in the final model.

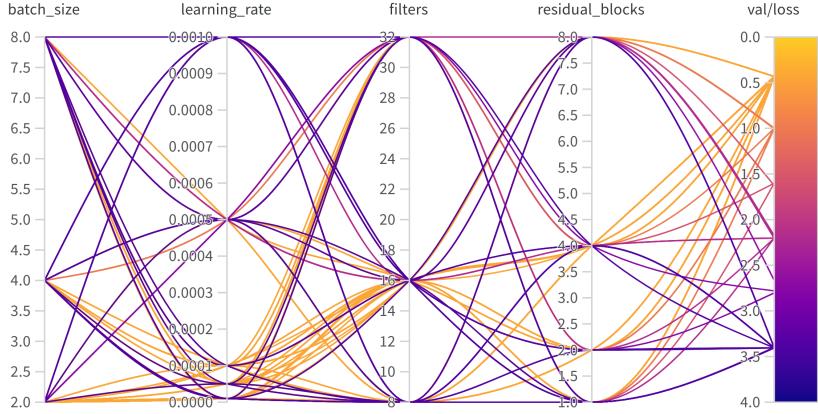


Figure 4.6: Influence of the model hyper-parameters on the validation loss (*val/loss*).

4.3.1 Model Hyper-Parameter Tuning

The model hyper-parameter-search aims to optimise parameters based on the model’s learning ability. For this sweep, we use a Poisson loss and square root data scaling since this resulted in desirable results in initial testing. The hyper-parameters we try and their range are:

- Amount of RRDB convolutional filters: [8, 16, 32]
- Amount of RRDB blocks: [2, 4, 8]
- Learning rate: [10^{-3} , 5×10^{-3} , 10^{-4} , 5×10^{-4} , 10^{-5} , 5×10^{-5}]
- Batch size: [2, 4, 8]

We trained models for every possible combination of these hyper-parameters (180 models) and monitor the loss of the validation data, see Figure 4.6. Several runs fail to converge. Some of these runs result in a validation loss greater or equal than 1.0, and only generate blank images.

Filtering out failed and poorly performing runs ($\text{val/loss} \geq 0.434$) still leaves us with a huge number of viable model hyper-parameter combinations (Figure 4.7). We, therefore, look at the correlation of the parameters with the loss.

The batch size has the largest positive correlation with the final loss. This positive correlation indicates that the bigger the *batch_size*, the worse the performance. After discarding runs that use a batch size of 8, we find that the next biggest correlation comes from the learning rate. It was clear that many of the failed runs were a result of high learning rates that make the training unstable but it’s also known that small learning rates risk getting stuck at local minima. We therefore discard the extreme learning rates $lr > 0.0001$ and $lr < 0.00005$.

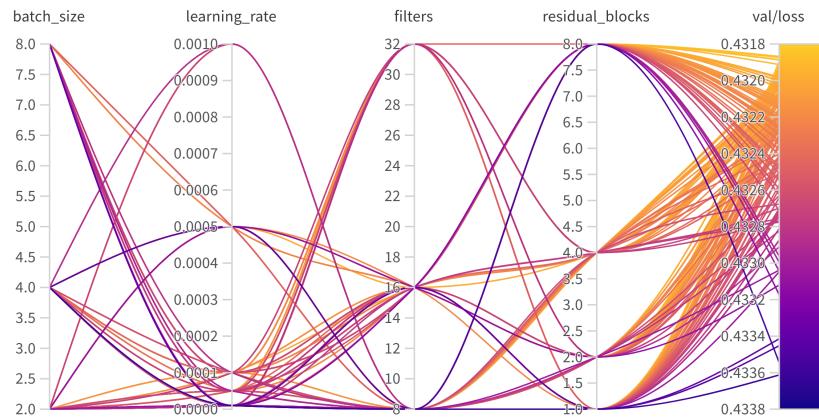


Figure 4.7: Influence of the model hyper-parameters on the validation loss ($val/loss$), with $val/loss \geq 0.434$.

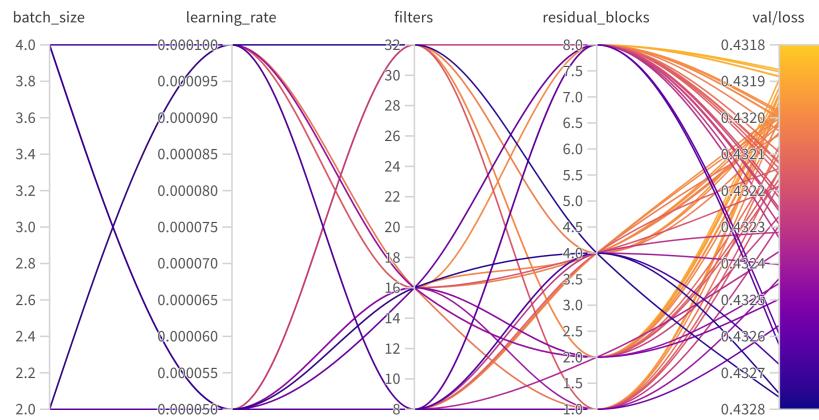


Figure 4.8: Influence of the model hyper-parameters on the validation loss ($val/loss$). With $val/loss \geq 0.434$, $batch_size \leq 4$ and $0.0001 \geq learning_rate \geq 0.00005$.

Figure 4.8 shows the remaining runs after filtering, all of which result in very similar validation loss values. Since these models are trained on a cropped image and reduced dataset size, we can assume that there is more to learn from the data and that a larger model would be more suitable for the real run. However, bigger models also take longer to train and use more GPU memory, which is a limiting factor when processing full-size images. For the final model (both SR and DN), we opt for $filters = 32$ and $num_residual_blocks = 4$ to leave room to learn more complex data without hitting our GPU memory limitation. We opt for a $batch_size = 1$ to similarly reduce the computational strain on the GPU and a learning rate of $lr = 0.0001$.

4.3.2 Data Hyper-Parameter Tuning

Having determined the model hyper-parameters we tune the hyper-parameters that influence the visual properties of the generated images: the loss function and data scaling. Here, we fix the model hyper-parameters to the optimal values determined in subsection 4.3.1, however, the batch size used is set to $batch_size = 4$ to increase the training speed. We train a model for all possible combination of loss functions (subsection 4.2.2) and data scalings (subsection 4.1.1, 128 models).

To determine the optimal data hyper-parameters we visually compare the generated images and their image quality metrics (subsection 4.2.3). Since we cannot consistently inspect the thousands of generated images, we first select the best performing models based on the evaluation metrics and before deciding the final data hyper-parameters based on both a quantitative and qualitative visual inspection.

Each evaluation metric emphasizes different visual elements in the generated image so we define our quantitative measure as a metric score that combines all the metrics into a single value. Some metrics are ascending, and others are descending so we invert ascending metrics such that all metrics are descending. Additionally, since the metrics map to different numerical scales, we apply a min-max normalization to their values before they are summed to create the combined metric score. Here, a lower combined metric score is better. Since images with different data scalings have different properties, the metrics' value is also differs.

Data Scale Determination

We correlate each hyper-parameter with the combined metric score (Figure 4.9) and find that the logarithmic scale correlates heavily with bad performance on all metrics. The asinh scale also under-performs with respect to the sqrt and linear scale. The sup-par performance of asinh and log is likely due to the noise level getting pushed close to the structure level, making it difficult to distinguish between the unpredictable background noise and any real features. The linear and sqrt data scalings perform the best.

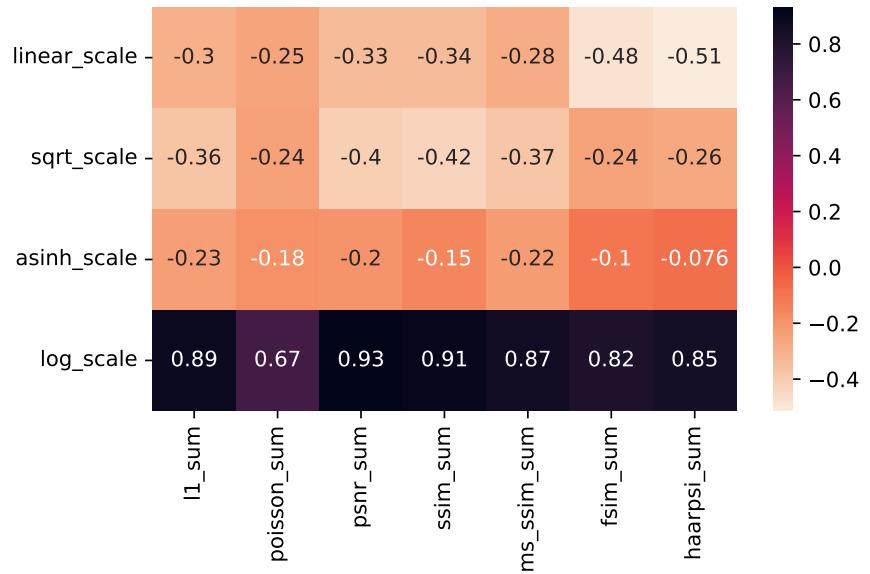


Figure 4.9: Correlation matrix data scaling with respect to image quality metrics. Lower correlation indicates better performance.

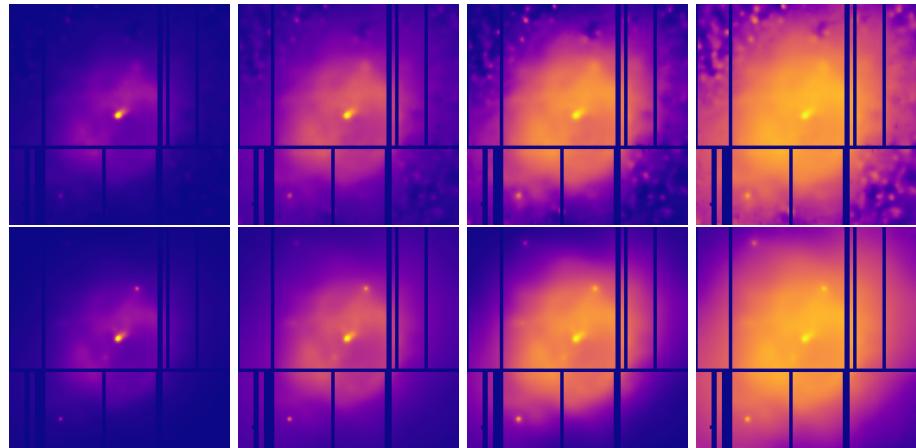


Figure 4.10: A model trained at linear data scale (top row) and a model trained at the sqrt data scale (bottom row). The display data scales from left to right: linear, sqrt, asinh, and log. At the asinh and log data scale, the *blobs* generated by the linear trained model are not present with the sqrt trained model.

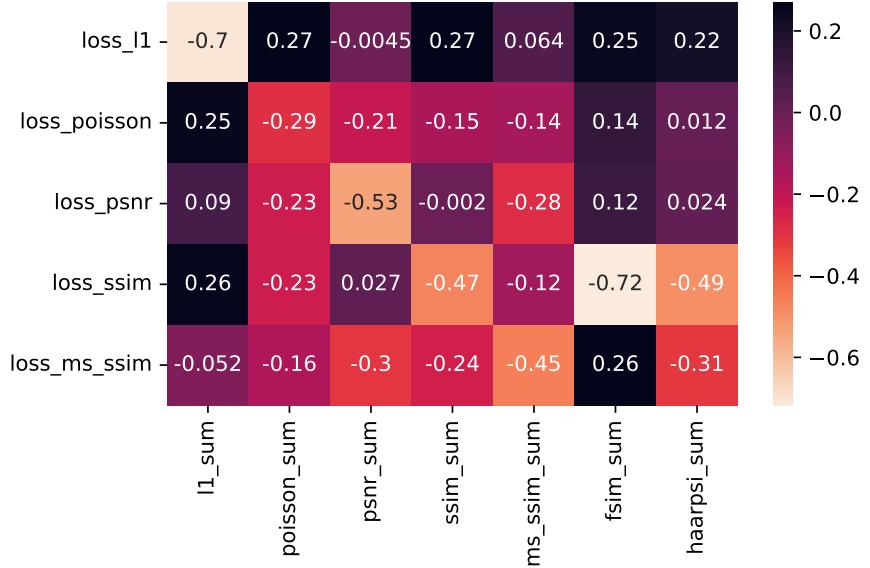


Figure 4.11: Correlation of the loss function elements with respect to the summed normalized metrics, given sqrt data scaling. Lower correlation indicates better performance.

Visual inspection of the linear data scaling models show the tendency of generating patchy images that are not present in the ground truth image. These artefacts are barely visible on linear scales, but they become problematic when the image is stretched. Models trained with a sqrt data scaling suffer less from this problem, see Figure 4.10, which motivates our choice of sqrt data scaling for our final model.

Loss Function Determination

Fixing the image scaling to sqrt, we then determine the optimal loss function. Again we correlate the loss function with respect to the combined metric score (Figure 4.11), and find that L1 loss only performs well with respect to the L1 evaluation metric and performs poorly with respect to all other metrics. As one might expect, the best correlations occur where the chosen loss function is also used as the evaluation metric. For the FSIM and HaarPSI evaluation metrics, the SSIM loss offers the best correlation.

When visually inspecting the generated images we observed that models trained with the SSIM loss tend to contain overly defined structures and AGNs in comparison to the target image. While models trained with Poisson, PSNR and MS_SSIM were visually closer to the target images. Models trained with L1 loss seem to suffer from a quantization problem, where there are distinct

RRDB convolutional filters	32
RRDB blocks	4
Batch size	1
Learning rate	0.0001
Data scaling	sqrt
Loss functions	SSIM, PSNR and MS_SSIM

Table 4.1: Final hyper-parameters

regions visible in what should be continuous distributed area. Based on these observations and the models performance on the image quality metrics we chose to train two final models with the loss functions:

- SSIM loss: The model trained with this loss performed the best on the SSIM, FSIM and HaarPSI metrics and visually generated the most defined structures and AGNs.
- PSNR combined with MS_SSIM loss: The model trained with this loss performed best on the PSNR and Poisson metrics and visually generated images similar to the target images. Since the PSNR (L_{psnr}) and MS_SSIM ($L_{\text{ms_ssim}}$) losses are at different scales we needed to normalize them in order to meaningfully combine them, as described in subsection 4.2.2. The final normalized combined loss (L_c) is defined as:

$$L_c = -0.0609 \cdot L_{\text{psnr}} - 1.51 \cdot L_{\text{ms_ssim}} + 5.43$$

The final hyper-parameters are shown in Table 4.1.

4.4 Final Model Training

The final models are trained on the full simulated training dataset. We train the models for 50 epochs using the final hyper-parameters (Table 4.1) and an Adam [60] optimiser with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. After the training is complete we select the model which achieved the best validation loss as our final model.

The input to the super resolution model is the simulated *XMM-Newton* image at the normal resolution with an exposure time of 20ks and background noise. For the target image we use the simulation image with 2x resolution, an exposure time of 100ks and without background noise. Omitting the background noise from the label image allows the model to concentrate on the source.

For the de-noising model, the input image is at the default *XMM-Newton* resolution with 20ks exposure time and the target image is similarly at the default *XMM-Newton* resolution but with 50ks exposure time. We choose this combination of exposure times to replicate more realistic observations and to ensure our results are trustworthy. Having an exposure time of for example 100ks would force the

model to make more uncertain predictions based on the input image. At 50ks exposure we also have more real world data to train and validate on.

Transfer learning

For the de-noising model we additionally have available real data. Whilst the simulated dataset contains more images, the real data encodes the domain that we are interested in. With the smaller dataset of real images, the performance of training a model on real data alone could be limited. Instead we make use of transfer learning [61] by taking the model trained on the larger simulated dataset and fine-tuning the weights to optimise for the real data. Fine-tuning is done by further training the model using the real data for another 50 epochs. We again select the model which performed best on the validation loss as our final model.

Chapter 5

Results

In this chapter the results of training the final super-resolution and de-noising models are presented. We first focus on the super-resolution models (section 5.1), which are analysed on the simulated data based on their metrical performance, visual analysis and brightness curves. Based on these results we select the final *XMM-SuperRes* model (subsection 5.1.3). To validate this model we will qualitatively compare the generated images with their *Chandra* counterparts, an X-ray telescope with a higher resolution.

Next, we will present the results of the training of the final de-noising models (section 5.2). The results of training on simulated data, real data and fine-tuning will be presented, from which the final *XMM-DeNoise* model will be selected. This model is finally compared to the non-ML wavelet transform method. All figures are with the square root function, unless stated otherwise, to visually enhance pixels with lower counts.

5.1 Super-Resolution Model

We trained the final SR models, one with SSIM loss and one with PSNR and MS_SSIM loss, as described in section 4.3. The performance of these models based on the simulated test set are shown in Table 5.1. The metrics are calculated compared to the target image using the unscaled (linear) data. To be able to do a comparison between the input and the target images, we need to match their resolutions. We use a naive method, namely nearest-neighbour up-sampling. The model trained with PSNR and MS_SSIM loss outperformed the model trained with SSIM loss on all metrics.

Metric	Input	SSIM Loss	PSNR and MS_SSIM Loss
L1	0.01096	0.006671	0.006508
PSNR	33.525	37.419	38.034
Poisson	0.08285	0.05014	0.04997
SSIM	0.8248	0.9065	0.907
MS_SSIM	0.9499	0.9833	0.9846
FSIM	0.8657	0.8631	0.8688
HaarPsi	0.5312	0.6833	0.697

Table 5.1: Super-resolution model image quality metrics based on the simulated test set. Green is the best performing score and orange the worst.

Name	Type	Comments
Bullet Cluster (RXJ0658-88)	Colliding clusters of galaxies	Clear shockwave
W49B (SNR G043.3-00.2)	Supernova remnant	Shockwaves and cavities
M51	Galaxy group	Many AGNs

Table 5.2: Real XMM-Newton EPIC-pn observation selection.

Name	Comments
TNG50-1 Subhalo 382215	Structure in combination with AGNs
TNG100-1 Subhalo 41583	Multiple distinct shockwaves, not many AGNs
TNG300-1 Subhalo 296363	Cavity with bright center source

Table 5.3: Simulated XMM-Newton EPIC-pn observations selection.

5.1.1 Visual Analysis

For the qualitative analysis of the model, we selected real and simulated observations. The selection of real observations is displayed in Table 5.2 and of simulated observations in Table 5.3. The observations cover a range of sources we are interested in, such as supernova remnants and galaxy clusters with different properties.

In Figure 5.1 we can see a few example generated super-resolution images. The images are cropped to highlight the details. For the uncropped full-frame images see Figure A.2. The input is a simulated XMM-Newton image at the normal resolution at 20ks with background noise. For the target image we use the simulation image with 2x resolution, an exposure time of 100ks and without background noise. Omitting the background noise from the label image allows the model to concentrate on the source. The generated images tend to contain more defined structures and more AGN. The model trained with the SSIM loss generates images with more contrast, but also tends to generate more *blobs* in for example shock waves. The model trained with PSNR and MS_SSIM loss generate images with similar contrast to the label images and less *blobs* in continuous regions such as shock waves. These *blobs* could be a bias for over predicting AGN, since AGN are abundant in our images.

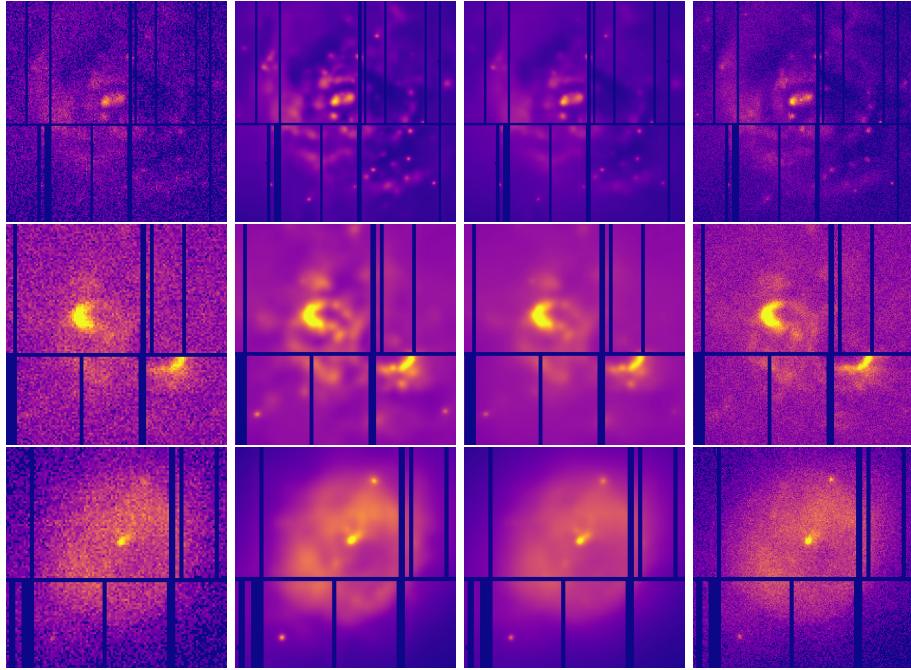


Figure 5.1: Super-resolution and de-noising examples from the simulated XMM dataset. Cropped to the central source and scaled with the square root function. From *top* to *bottom*: TNG50 Subhalo 382215, TNG100 Subhalo 41583 and TNG300 Subhalo 296363. From *left* to *right*: Input image at 1x resolution with 20ks exposure, generated super-resolution image for 100ks with the SSIM loss model, generated super-resolution image for 100ks with the PSNR combined with MS_SSIM loss model and the label image at 100ks without background noise.

5.1.2 Brightness Analysis

To analyze the performance SR and DN models in more detail, we take vertical and horizontal segments of the input and generated images (Figure 5.2) and plot the pixel value distribution summed along the minor axis (Figure 5.3). The generated images are smoother than the input and target images; therefore, we smooth the input and target images using 1d convolution with a Gaussian kernel of size 5 and $\sigma = 1.0$ for a fairer comparison. Since the input image is at 20ks and the target image is at 100ks, the input image will have lower counts per region. The sudden drop in count values corresponds with the chip-gaps where the count is zero. The model trained with SSIM loss (red) has more contrast - it has higher signal peaks and lower lows than the target image. The model trained with PSNR and MS_SSIM loss (magenta) closely resembles the target brightness.

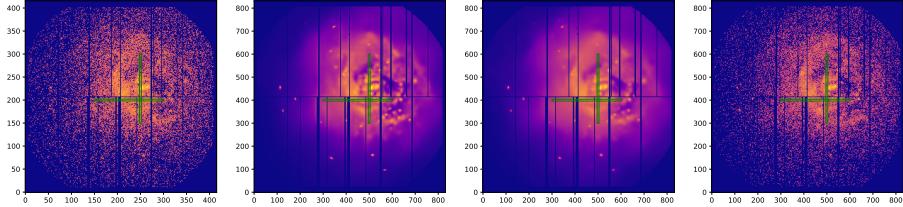


Figure 5.2: Strip plot regions of the simulated source TNG50 Subhalo 382215, scaled with a logarithmic function. From left to right: input image at 1x resolution at 20ks, generated image with SSIM loss trained model with 2x resolution at 100ks, generated image with PSNR combined with MS_SSIM loss trained model with 2x resolution at 100ks and the target image with 2x resolution at 100ks without background noise. The green regions indicate the regions that we will analyse on brightness.

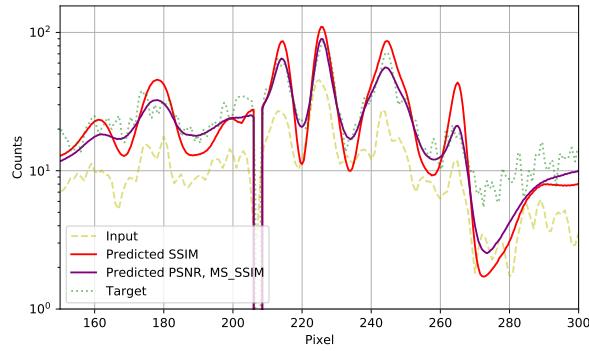
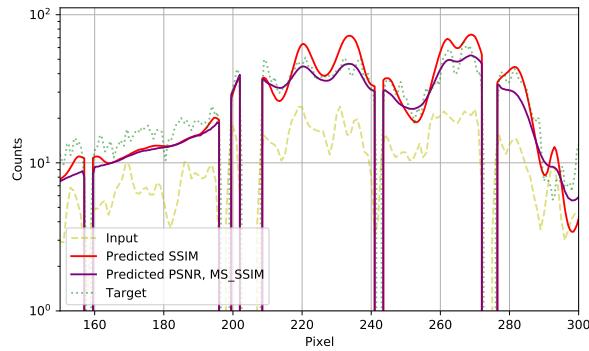


Figure 5.3: Count plots of the horizontal cutouts (top) and vertical cutouts (bottom) of the simulated source TNG50 Subhalo 382215.

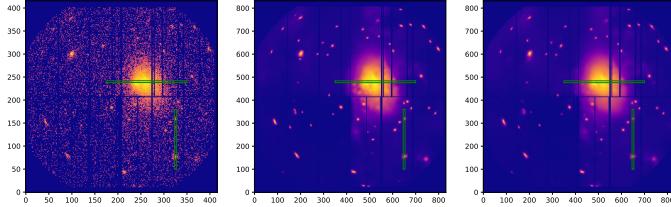


Figure 5.4: Strip plot regions an *XMM-Newton* observation of the Bullet Cluster (RXJ0658-88, obsid 0112980201), scaled with a logarithmic function. From left to right: input image at 1x resolution at 20ks, generated image with SSIM loss trained model with 2x resolution at 100ks and the generated image with PSNR combined with MS_SSIM loss trained model with 2x resolution at 100ks. The green regions indicate the regions that we will analyse on brightness.

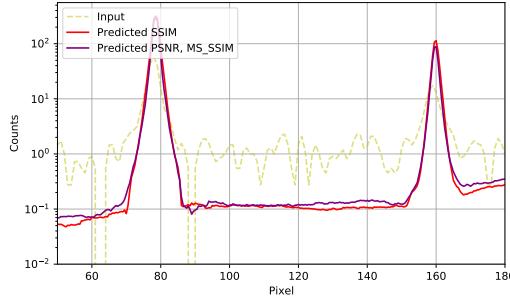
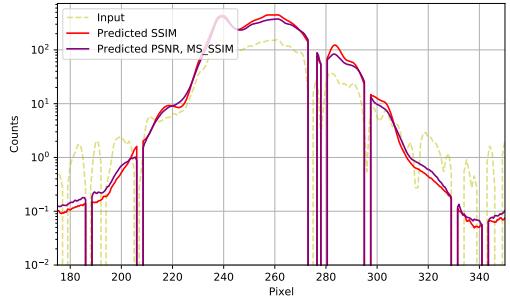


Figure 5.5: Count plots of the horizontal cutouts (top) and vertical cutouts (bottom) of the Bullet Cluster.

We also took vertical and horizontal cutouts on the real *XMM-Newton* observation of Bullet Cluster (RXJ0658-88, obsid 0112980201) (Figure 5.4). The corresponding pixel value distribution (Figure 5.5) showcases that point sources are much better defined in the generated images. In between the point sources the noise level is also lower.

5.1.3 XMM-SuperRes Model

We select the SR and DN model trained with PSNR and MS_SSIM loss for our final model, named *XMM-SuperRes*. Because this outperforms the model trained with SSIM loss metrically, is closer to the target images in the visual inspection and generates flux intensities closer to the target images. Further analysis will only be done this *XMM-SuperRes* model.

5.1.4 Chandra Comparison

For the super-resolution model, it is not possible to learn the domain mapping for real *XMM-Newton* observations. However, we can probe the performance of the super resolution model applied to the real data by comparing the generated images with their *Chandra* counterparts.

As a qualitative measure of our results we compare both our SR and DN *XMM-Newton* generated images with *Chandra* observations of the same source. The *Chandra* images we compare against have a higher resolution of 0.5 arcsec FWHM compared to the 6 arcsec FWHM of the *XMM-Newton* EPIC-pn. We use the full exposure time of the *Chandra* images. We do however stress that the properties of the 2 telescopes are not equivalent. The PSF of the two instruments are not the same and *XMM-Newton* is more sensitive than *Chandra* so these images can not be considered as a ground-truth. None the less we present a select few study cases to cover a variety of structures.

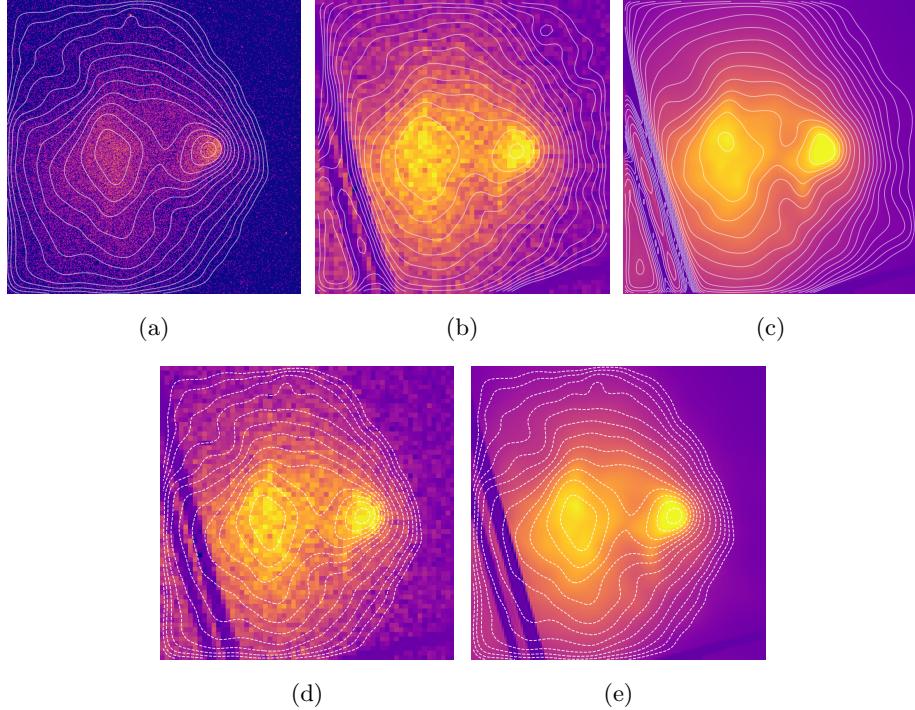


Figure 5.6: Images of the two colliding clusters of galaxies Bullet Cluster (1E 0657-56) with contours highlighted in white. With: *Chandra* at 88ks exposure (a), *XMM-Newton* at 20ks exposure (b), generated *XMM-Newton* SR and DN (c), *XMM-Newton* at 20ks exposure overlayed with the *Chandra* contours (d) and generated *XMM-Newton* SR and DN overlayed with the *Chandra* contours (e).

Our first case study is the Bullet cluster (Figure 5.6) that depicts two colliding clusters of galaxies. The cavity between the two galaxies is enhanced in both the *Chandra* (a) and the generated SR and DN image (c) in comparison to the input *XMM-Newton* image (b). Looking at the real *XMM-Newton* image (d) and generated SR and DN *XMM-Newton* (e) with in white the contours of *Chandra* overlaid. We can see that the cavity between the two clusters is much better defined in the SR and DN image compared to the original *XMM-Newton* observation, the *Chandra* observation also clearly contains this source.

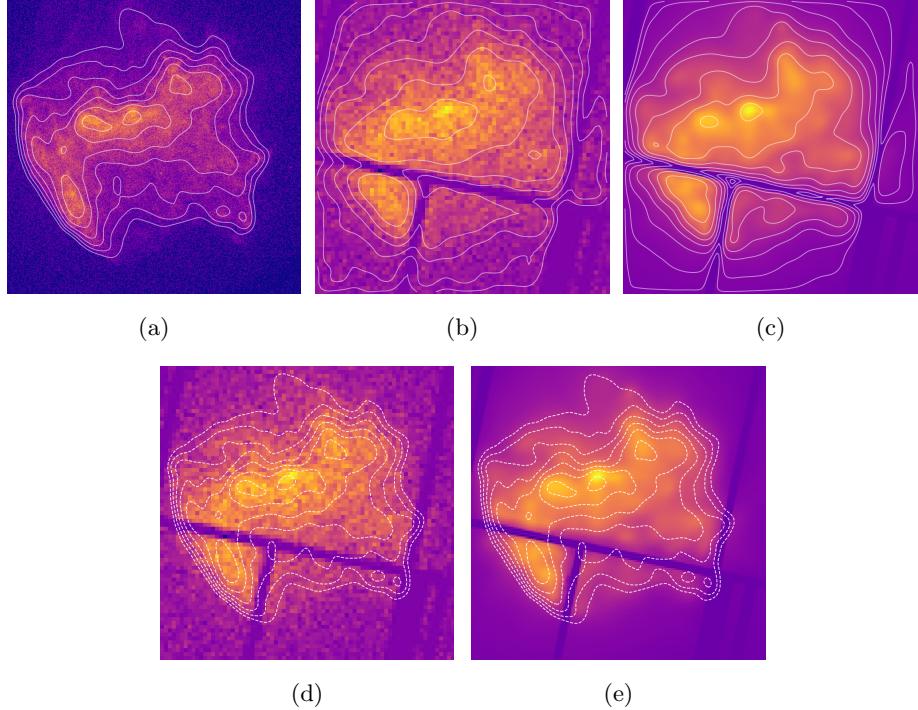


Figure 5.7: Images of the supernova remnant W49B (SNR G043.3-00.2) with contours highlighted in white. With: *Chandra* at 88ks exposure (a), *XMM-Newton* at 20ks exposure (b), generated *XMM-Newton* SR and DN (c), *XMM-Newton* at 20ks exposure overlayed with the *Chandra* contours (d) and generated *XMM-Newton* SR and DN overlayed with the *Chandra* contours (e).

Our next case is supernova remnant W49B (Figure 5.7). Here, we again see the more pronounced features in the SR+DN image (c) in comparison to the input (b). The extended features on the top of the image seen in the *Chandra* contour lines are better defined in the generated SR and DN image (e) compared with the real *XMM-Newton* image (d).

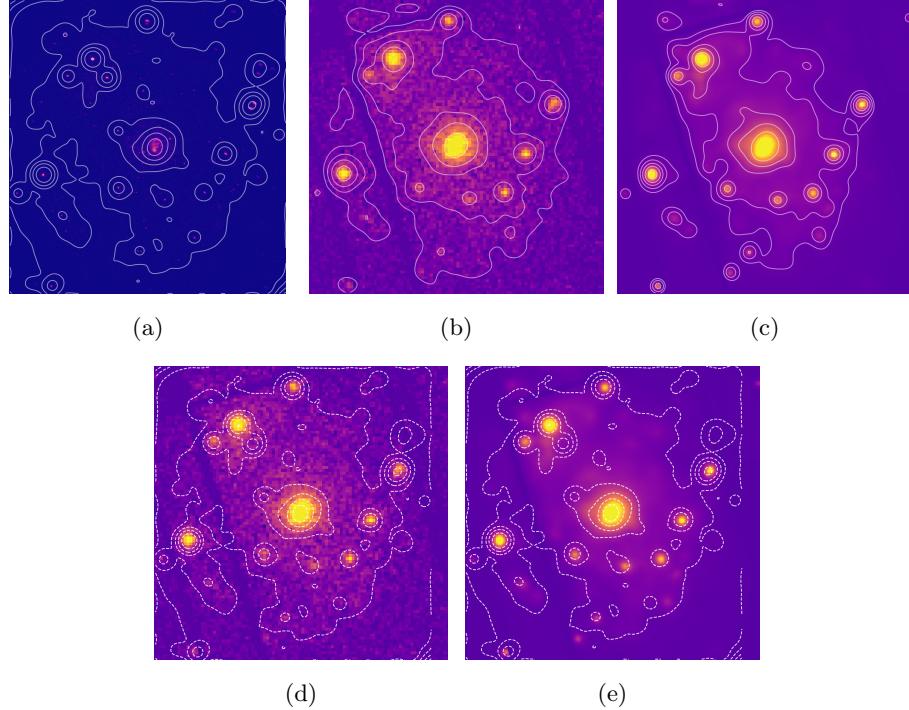


Figure 5.8: Images of the group of galaxies M51 with with contours highlighted in white. With: *Chandra* at 88ks exposure (a), *XMM-Newton* at 20ks exposure (b), generated *XMM-Newton* SR and DN (c), *XMM-Newton* at 20ks exposure overlayed with the *Chandra* contours (d) and generated *XMM-Newton* SR and DN overlayed with the *Chandra* contours (e).

The group of galaxies M51 is nice case study to see how the network performs more compact sources (Figure 5.8). In this example we see that the generated image (c) has AGNs (point sources) that are better defined compared to the real *XMM-Newton* image (b). For example, the faint AGN at the bottom left of the AGN is clearly visible in the SR and DN (c) and *Chandra* image (a) but are barely in the real *XMM-Newton* image (b). Looking at the real *XMM-Newton* image (d) and generated SR and DN *XMM-Newton* (e) with in white the contours of *Chandra* overlaid. We can also see that in the top left of the SR and DN image (e) an extended feature is visible that matches with the contours of *Chandra*, this extended feature is harder to be seen in the real *XMM-Newton* image (d). However, the SR and DN image also does sometimes mis-predict features, for example on the right side there is a circular *blob* visible in the *Chandra* contours. In the real *XMM-Newton* image it is hard to tell if there is anything present. However, the SR model predicted almost no counts in that area.

Metric	Input	simulated data, SSIM loss	simulated data, PSNR and MS_SSIM loss	real data, SSIM loss	real data, PSNR and MS_SSIM loss	fine-tuned, SSIM loss	fine-tuned, PSNR and MS_SSIM loss
L1	0.0065	0.005045	0.005202	0.005127	0.004628	0.005045	0.004408
PSNR	39.349	41.608	41.728	41.699	42.227	42.594	42.693
Poisson	0.07616	0.04785	0.04782	0.04919	0.04856	0.04781	0.04778
SSIM	0.9484	0.942	0.9359	0.9613	0.9512	0.9571	0.9567
MS_SSIM	0.9922	0.9916	0.991	0.9941	0.993	0.9938	0.9939
FSIM	0.9688	0.9611	0.9577	0.9758	0.9745	0.9782	0.9783
HaarPsi	0.8879	0.8888	0.9006	0.9095	0.9139	0.9232	0.9253

Table 5.4: De-Noising model metric score on the test set without data scaling, input at 20ks exposure and the target at 50ks exposure. The orange cells indicate the worst performance and green the best performance per metric.

5.2 De-Noising Model

In Table 5.4 the performance of the de-noising models trained to convert 20ks exposure images to 50ks exposure images based on the real data test set without data scaling is displayed. We can see that except for the model trained on simulated data with the PSNR and MS_SSIM loss, all the models improved compared to the input. The models trained on the real dataset generally score better than those trained on the simulated data. This could be explained by the fact that the test set was based on real data and that certain features present in the real data are less present in simulated data. The model that performed the best overall is the model with PSNR and MS_SSIM loss that was first trained on simulated data and then fine-tuned on real data. Only being outperformed on the SSIM and MS_SSIM metrics, this is not completely surprising considering that the other model was trained using only the SSIM loss. We, therefore, select the DN model trained with PSNR and MS_SSIM loss for our final model, named *XMM-DeNoise*. Further analysis will only be done using this model.

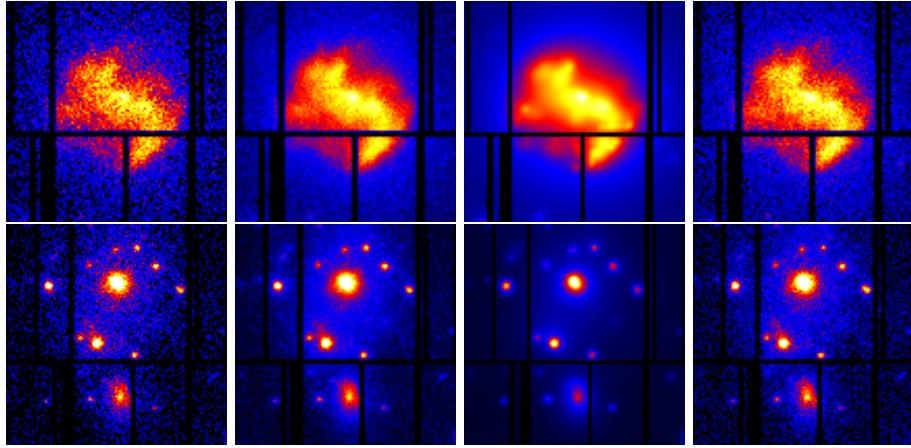


Figure 5.9: De-noised and wavelet transformed examples, W49B (top) and M51 (bottom), from the real XMM dataset. Cropped to the central source and scaled with the square root function. From *left* to *right*: Input image at 1x resolution with 20ks exposure, generated de-noised image for 50ks, wavelet transformed image and the target image at 50ks.

5.2.1 Wavelet Comparison

We qualitatively compare our *XMM-Denoise* model to the non machine learning based wavelet transform. The use of wavelet based de-noising methods has been shown to optimize the detection of AGNs, galaxy clusters and other features in X-ray images of different telescopes [34][35][36]. Our implementation is based on [34].

In Figure 5.9 de-noised examples generated by *XMM-DeNoise* are shown compared to wavelet transformed image and the target image. The images are cropped to highlight the details. For the uncropped full-frame images see Figure A.4. We can see that our de-noised images are, compared to the more smoothed wavelet transformed images, visually much closer to the target images. Note that for the wavelet technique the goal is not to mimic the higher exposure time image but to de-noise the images substantially. Certain features, such as the shock waves in W49B (top row) are better defined because of this. However, this also comes at the risk of having more artifacts or filtering out too much information. For example, in the M51 images (bottom row) we can see that the wavelet transformations filtered out the extended features of the source in the center left of the image. This is a known limitation of using wavelet transforms for non spherical morphologies.

Chapter 6

Discussion

This chapter will discuss the experimental results to answer our initial research questions. We will also discuss interesting observations resulting from our experiments, note limitations, and propose suggestions for future work.

6.1 Research Questions

The in-depth research for this thesis started by focusing on four research questions. In this section, we will discuss the experimental results to answer these research questions.

Research Question 1: How can we use GAN models to increase the resolution of XMM-Newton X-ray images and increase their scientific value?

For the initial SR prototypes, we took an approach based on GAN models. We experimented with different GAN-based architectures and found that a model based on ESR-GAN [19] worked well on generating realistic-looking SR and DN images. However, as explained in section 4.2, ESR-GAN suffers from hallucinations of non-existent features in the model output. This is likely caused by the adversarial loss, which helps to generate images visually indistinguishable when compared to high-resolution target images. However, in order to do this, the adversarial loss might push the model to generate realistic looking features that do not actually exist in the target image that we are trying to predict. This is problematic for our use-case. Removing the adversarial loss greatly reduced this problem. However, this new model is only able to generate sharp reconstructions in areas of high confidence. Aspects of low confidence such as the background noise will result in blurry reconstructions and the output image is unlikely to visually resemble the target. We argue that for the scientific value, it is preferable to have a more “*blurry*” but more trustworthy image. We, therefore, decided not to continue with GAN models. The generation of realistic-looking images of the GAN model did demonstrate the effectiveness of

the RRDB based generator model used in the ESR-GAN architecture. We therefore decided to continue the research using this RRDB generator architecture. Based on this architecture, we developed our *XMM-SuperRes* model, which is able to generate SR and DN images from real *XMM-Newton* observations. In these SR and DN images (extended) features such as shock waves, cavities and AGNs are often better perceivable than in the original image. This could help astronomers find interesting properties of sources in *XMM-Newton* observations, increasing their scientific value.

Research Question 2: How can we use U-net CNN-based models to de-noise XMM-Newton X-ray images and increase their scientific value?

With a few adaptations, we made the RRDB generator architecture also suitable and effective for de-noising. Using the same RRDB architecture had the additional benefit that key findings in the development of the *XMM-SuperRes* model were largely applicable to the de-noising model. Our *XMM-DeNoise* model de-noises *XMM-Newton* observations by generating images with 2.5 times higher exposure time. Increasing the exposure time is the primary way to increase the signal-to-noise ratio in real observations. The *XMM-DeNoise* model improves the image quality based on all the metrics we measured. Since this model was effective at de-noising, we decided that there was no need to research U-net CNN-based models. This also helped limit the scope of this research.

We found that training the de-noising model on simulated data first and then fine-tuning it on real data resulted in the best results for most image quality metrics. Another advantage of de-noising is that we trained and tested the *XMM-DeNoise* model on real data. Therefore, we did not have to worry about simulation inaccuracies and their resulting artifacts, something which does play a role with the *XMM-SuperRes* model. Also, the numerical image quality metrics are more meaningful on real data.

The *XMM-DeNoise* model tries to achieve something easier in comparison with the *XMM-SuperRes* model. With DN, we aim to predict an image of 2.5 times higher exposure. In comparison, the SR and DN model aims to predict 5 times higher exposure without additional background noise and twice the resolution. The easier problem it tries to solve makes the generated DN images more trustworthy since there is less opportunity for problematic spurious predictions. However, the potential insights from DN images are also less compared to the SR and DN images. Since with de-noising we do not change the fundamental image properties, such as the resolution, existing *XMM-Newton* analysis tools can be used more easily. De-noised images could help astronomers find interesting properties of sources in *XMM-Newton* observations while using lower exposure time observations. There are exponentially more existing observations at lower exposure times, increasing their scientific value. Being able to use lower exposure time observations would also enable the *XMM-Newton* to do more observations.

Note that due to the nature of the reconstruction, features in the generated

DN image might be spurious. Therefore, to validate the discoveries made in the enhanced images, further detailed analysis and ideally follow-up observations at higher spatial resolution will be needed for their confirmation.

Overall, this would enable the *XMM-Newton* telescope to do more observations, increasing its scientific value.

Research Question 3: How does the scientific reliability of the improved images compare with published results using conventional, non-ML methods?

There is no metric for scientific reliability, but based on the qualitative comparisons of the SR and DN images generated with *XMM-SuperRes* model, we find that the enhanced features have obvious counterparts in the higher resolution *Chandra* observations. We also showed that the enhanced simulated images generated by *XMM-SuperRes* tend to follow the brightness of the target images closely. Unfortunately, we were not able to successfully implement direct demodulation method for this research. We, therefore, have no direct comparison between our *XMM-SuperRes* model and a conventional super resolving method.

We validated the performance of our *XMM-DeNoise* model on real data. This makes the image quality metrics more meaningful since the real data encodes the domain of interest. We found that it improved the PSNR by 8.2% and reduced the L1 by 38%. Compared to the non-ML wavelet transformed images, our generated DN images are visually closer to the target images, whereas the wavelet transformed image produces a more smoothed image. The advantage of the wavelet transformed images is that it tends to highlight certain features more clearly. However, this also comes at the risk of having more artifacts or filtering out too much information. We also found that our method works better for nonspherical morphologies. Finally, we argue that since our enhanced images are visually closer to the target images and contain less artifacts compared wavelet transformed images, they are scientifically more reliable.

Research Question 4: If the scientific validation from point 3 is successful, can we obtain new results on potentially interesting targets and publish them?

At the time of writing, no research has been published using deep-learning-based SR and DN methods for X-ray astronomy. We think that the insights obtained in this thesis are beneficial in furthering the field of X-ray astronomy. Therefore, we are currently in the process of writing a scientific paper with the intent to publish in the *Monthly Notices of the Royal Astronomical Society*.

6.2 Limitations and Future Work

6.2.1 Reliability

Due to the nature of deep-learning-based SR and DN, the model can make wrong predictions. In this thesis, we propose multiple methods to address this problem and qualitatively show that the model generated images that are able to enhance features with obvious counterparts in the *Chandra* observations, which has a higher spatial resolution. Nevertheless, due to the nature of the reconstruction, some of the generated features may be spurious; hence whilst this model may be able to find and uncover interesting details, further detailed analysis and ideally follow-up observations at higher spatial resolution will be needed for their confirmation.

The root of the reliability problem stems from the fact that SR and DN are inherently an ill-posed problem. Given a low-resolution or noisy image, multiple SR or DN images are possible, as explained in section 4.2. Our current SR and DN model only predicts one of the possible output images. There are SR and DN architectures that can output a range of SR and DN images, such as SRFflow [21]. Using the SRFflow architecture, we can set a parameter at inference that changes the output properties of the images. Setting this parameter to a low value results in generated images with minimal artifacts but that are very “*blurry*” and do not generally contain more information than can be found in the input image. Setting it to a high value, more realistic images are generated, but they tend to contain more artifacts. This parameter can be set by the user depending on the desired output. We experimented with this architecture, and the initial results looked promising. However, to limit the scope of this research, we decided not to continue with it for this version. This could be an interesting starting point for future research.

6.2.2 Bright Sources

The range of pixel values in *XMM-Newton* observations can be very large. A bright AGN might cause thousands of counts in one pixel, but a shock-wave of a supernova remnant may only result in a few counts in another pixel. We found that this extremely dynamic range was problematic for the training stability of our models. One method to combat this problem was to clip the extreme values. Unfortunately, this clipping will cause information to be lost in bright sources. In section A.9 an example is illustrated. Since our main objective is to improve the resolution and de-noise so that extended structures are more clearly visible, losing detail in bright sources seems an acceptable compromise. However, when analyzing sources to compare with existing literature, we found that some of these sources have bright parts above the clipping limit. The clipping limit could be increased to make the models more usable on these brighter sources. Note that with a higher clipping limit, the insights found on which data scaling function to use and possibly the loss function might change.

6.2.3 Usable Domain

The current models are limited to the energy range $[0.5, 2]$ keV. We chose this energy range since most of the emission of extended sources happen within this range. Possibly the models could be applied to images in different energy bands. However, some complications might arise, such that the vignetting and noise properties are energy-dependent. The PSF is less dependent on the energy range. Also, both our models take 20 ks exposure time observations as input. We chose this exposure time since it seems a good trade-off between having many observations but still containing enough information to generate interesting results. We only focus on the *XMM-Newton* EPIC-pn sensor since it has the largest effective area and, therefore, good spectral and spatial resolutions. We argue that the domain chosen for this research is sufficient to show the effectiveness of our proof-of-concept method. Future research could look into training a model that accepts wider energy ranges, incorporates more sensors, such as the MOS detectors, and accepts any exposure time by, for example, passing this as an input parameter.

6.2.4 Simulation Limitations

In order to train the SR and DN models, we created a custom *XMM-Newton* EPIC-pn simulator to generate a training dataset. We decided not to simulate some telescope properties such as out-of-time events to simplify and speed up the simulator. The sources we are interested in for this research are usually not extremely bright, which causes out-of-time events. When the model enhances a bright source it tends to produce many artifacts, as can be seen in Figure A.6. For future research, it could be interesting to extend the models to be able to analyze brighter sources. In this case, the simulator could be extended to consider more of these properties. Also, the simulator could be extended to incorporate more *XMM-Newton* sensors.

6.2.5 AGN Deblending

This research primarily focused on extended sources and less on AGNs. However, when analyzing the results, the models seem to be good at enhancing faint AGNs or even deblending two AGNs. Deblending is of use when two AGNs are so close together that the telescope resolution is too low to separate them, i.e. they are perceived as one. However, if the telescope's resolution was higher, the AGNs would show up as two distinct point sources. The process of deblending finds and separates AGNs that are this close together. Future research could focus on using the *XMM-SuperRes* model for this purpose or even training a new SR and DN model specialized in deblending.

Chapter 7

Conclusion

In this thesis we have developed deep-learning-based super-resolution (SR) and de-noising (DN) models to enhance astronomical *XMM-Newton* X-ray EPIC-pn images. We increase the resolution of the observations and de-noise to improve the noise properties and enhance features that are challenging to locate in the original *XMM-Newton* images.

To train the SR and DN models, we simulated *XMM-Newton* EPIC-pn images with twice the nominal spatial resolution and images with larger exposure times. We explored the influence of the model architecture parameters, data pre-processing, and loss functions on the model’s performance. To enhance the image quality, we proposed using a combined loss function consisting of both PSRN and MS.SSIM. To address the problem of the high dynamical range of pixel values present in X-ray images, we implemented data scaling. We showed that using suitable data scaling, our models generated fewer artifacts in low surface brightness areas in extended sources while preserving the “interesting” features.

Our SR and DN model (*XMM-SuperRes*) generates enhanced SR and DN images with twice the spatial resolution and an improved PSNR. The network-produced images have the desired properties, such as a smaller PSF, and all the tested image quality metrics were improved when the model was applied on the test dataset. Specifically, it improved the PSNR by 21.5% and reduced the L1 by 40.3% when tested on simulated data. We validate the performance of the model applied on real data by comparing with observations taken by NASA’s *Chandra* telescope, which has much higher spatial resolution.

We find that the model generated images that are able to enhance features with obvious counterparts in the *Chandra* observations. Nevertheless, due to the nature of the reconstruction, some of the generated SR features may be spurious; hence whilst this model may be able to find and uncover interesting details, further detailed analysis and ideally follow-up observations at higher spatial resolution will be needed for their confirmation.

Our de-noising model (*XMM-DeNoise*) based on *XMM-SuperRes*, generates images with 2.5 times higher exposure without increasing the resolution.

This enabled us to train and validate the model on real *XMM-Newton* observations. We found that training the de-noising model on simulated data first and fine-tuning it on real data resulted in the best results for most image quality metrics. *XMM-DeNoise* similarly improves the quality of real *XMM-Newton* observations on all measured image quality metrics. Specifically, it improved the PSNR by 8.2% and reduced the L1 by 38%. Compared to wavelet transformed counterparts, we found that de-noised images generated by our model visually better resemble the higher exposure target images, perform better on nonspherical morphologies and produce less artifacts.

In conclusion, we have demonstrated the feasibility of using deep-learning models to improve the spatial resolution and de-noising of *XMM-Newton* EPIC-pn X-ray astronomy images to increase their scientific value. The *XMM-SuperRes* and *XMM-DeNoise* models developed in this paper could be used as a proof-of-concept to create more elaborated methods. This could include creating a model that can output a range of possible SR and DN images to more directly tackle the ill-posed nature of the problem. In the future, the training of the models could be extended to more instruments, more X-ray telescopes, and incorporating more energy ranges and exposure times.

The use of deep-learning-based image enhancement will allow astronomical X-ray images to improve our understanding our X-ray sky and the workings of our universe.

Chapter 8

Acknowledgements

I would like to thank my supervisors at ESAC, ESA: Ivan Valtchanov, Maggie Lieu, Eva Verdugo, Antónia Vojtekova and Maria Santos-Lieo. For patience, guidance and support. They helped me work with astronomical X-ray data, and gave invaluable advice at every step of this research. I would like to thank my supervisors from Maastricht University, Dr. A. Briassouli and Dr. D.H. Cámpora Pérez for their feedback and helping me putting the results on paper. I would like to thank Florian Pacaud, from the Argelander Institut für Astronomie (AIFA), for helping with creating the simulator PSF files and generating the wavelet transformed images. Finally, I would like to thank friends and family for proof-reading and supporting me through every part of this research.

S.F. Sweere acknowledges support from the ESA traineeship program and ESA through the Faculty of the European Space Astronomy Centre (ESAC) - Funding reference ESA-SCI-SC-LE-033.

The IllustrisTNG simulations were undertaken with compute time awarded by the Gauss Centre for Supercomputing (GCS) under GCS Large-Scale Projects GCS-ILLU and GCS-DWAR on the GCS share of the supercomputer Hazel Hen at the High Performance Computing Center Stuttgart (HLRS), as well as on the machines of the Max Planck Computing and Data Facility (MPCDF) in Garching, Germany.

Bibliography

- [1] F. Jansen, D. Lumb, B. Altieri, J. Clavel, M. Ehle, C. Erd, C. Gabriel, M. Guainazzi, P. Gondoin, R. Much, *et al.*, “Xmm-newton observatory-i. the spacecraft and operations,” *Astronomy & Astrophysics*, vol. 365, no. 1, pp. L1–L6, 2001.
- [2] M. Santos-Lleo, N. ScharTEL, H. Tananbaum, W. Tucker, and M. C. Weisskopf, “The first decade of science with chandra and xmm-newton,” *Nature*, vol. 462, no. 7276, pp. 997–1004, 2009.
- [3] D. Wilkins, L. Gallo, E. Costantini, W. Brandt, and R. Blandford, “Light bending and x-ray echoes from behind a supermassive black hole,” *Nature*, vol. 595, no. 7869, pp. 657–660, 2021.
- [4] M. J. Turner, A. Abbey, M. Arnaud, M. Balasini, M. Barbera, E. Belsole, P. Bennie, J. Bernard, G. Bignami, M. Boer, *et al.*, “The european photon imaging camera on xmm-newton: the mos cameras,” *Astronomy & Astrophysics*, vol. 365, no. 1, pp. L27–L35, 2001.
- [5] L. Strüder, U. Briel, K. Dennerl, R. Hartmann, E. Kendziorra, N. Meidinger, E. Pfeffermann, C. Reppin, B. Aschenbach, W. Bornemann, *et al.*, “The european photon imaging camera on xmm-newton: the pn-ccd camera,” *Astronomy & Astrophysics*, vol. 365, no. 1, pp. L18–L26, 2001.
- [6] M. C. Weisskopf, H. D. Tananbaum, L. P. Van Speybroeck, and S. L. O’Dell, “Chandra x-ray observatory (cxo): overview,” in *X-Ray Optics, Instruments, and Missions III*, vol. 4012, pp. 2–16, International Society for Optics and Photonics, 2000.
- [7] A. Vojtekova, M. Lieu, I. Valtchanov, B. Altieri, L. Old, Q. Chen, and F. Hroch, “Learning to denoise astronomical images with u-nets,” *Monthly Notices of the Royal Astronomical Society*, 2020.
- [8] J. Sanders and A. Fabian, “Adaptive binning of x-ray galaxy cluster images,” *Monthly Notices of the Royal Astronomical Society*, vol. 325, no. 1, pp. 178–186, 2001.

- [9] H. Bourdin, E. Slezak, A. Bijaoui, and M. Arnaud, “A multiscale regularized restoration algorithm for xmm-newton data,” *arXiv preprint astro-ph/0106138*, 2001.
- [10] J.-L. Starck, E. Pantin, and F. Murtagh, “Deconvolution in astronomy: A review,” *Publications of the Astronomical Society of the Pacific*, vol. 114, no. 800, p. 1051, 2002.
- [11] K. G. Puschmann and F. Kneer, “On super-resolution in astronomical imaging,” *Astronomy & Astrophysics*, vol. 436, no. 1, pp. 373–378, 2005.
- [12] Z. Li, Q. Peng, B. Bhanu, Q. Zhang, and H. He, “Super resolution for astronomical observations,” *Astrophysics and Space Science*, vol. 363, no. 5, pp. 1–15, 2018.
- [13] F. Zhou, W. Yang, and Q. Liao, “Interpolation-based image super-resolution using multisurface fitting,” *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3312–3318, 2012.
- [14] W.-C. Siu and K.-W. Hung, “Review of image interpolation and super-resolution,” in *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1–10, IEEE, 2012.
- [15] H. Feng, Y. Chen, S. Zhang, F. Lu, and T. Li, “Improving the spatial resolution of xmm-newton epic images by direct demodulation technique,” *Astronomy & Astrophysics*, vol. 402, no. 3, pp. 1151–1155, 2003.
- [16] W. Yang, X. Zhang, Y. Tian, W. Wang, J.-H. Xue, and Q. Liao, “Deep learning for single image super-resolution: A brief review,” *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3106–3121, 2019.
- [17] X. Zhang, Q. Chen, R. Ng, and V. Koltun, “Zoom to learn, learn to zoom,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3762–3770, 2019.
- [18] Z. Wang, J. Chen, and S. C. Hoi, “Deep learning for image super-resolution: A survey,” *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- [19] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, “Esrgan: Enhanced super-resolution generative adversarial networks,” in *Proceedings of the European conference on computer vision (ECCV) workshops*, pp. 0–0, 2018.
- [20] C. Chen, Q. Chen, J. Xu, and V. Koltun, “Learning to see in the dark,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3291–3300, 2018.

- [21] A. Lugmayr, M. Danelljan, L. Van Gool, and R. Timofte, “Srfflow: Learning the super-resolution space with normalizing flow,” in *European Conference on Computer Vision*, pp. 715–732, Springer, 2020.
- [22] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *European conference on computer vision*, pp. 184–199, Springer, 2014.
- [23] V. Jain and S. Seung, “Natural image denoising with convolutional networks,” *Advances in neural information processing systems*, vol. 21, 2008.
- [24] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [25] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *European conference on computer vision*, pp. 694–711, Springer, 2016.
- [26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [27] K. Schawinski, C. Zhang, H. Zhang, L. Fowler, and G. K. Santhanam, “Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit,” *Monthly Notices of the Royal Astronomical Society: Letters*, p. slx008, Jan 2017.
- [28] L. Lauritsen, H. Dickinson, J. Bromley, S. Serjeant, C.-F. Lim, Z.-K. Gao, and W.-H. Wang, “Super-resolving herschel imaging: a proof of concept using deep neural networks,” *arXiv preprint arXiv:2102.06222*, 2021.
- [29] L. Strüder, U. Briel, K. Dennerl, R. Hartmann, E. Kendziorra, N. Meidinger, E. Pfeffermann, C. Reppin, B. Aschenbach, W. Bornemann, *et al.*, “The european photon imaging camera on xmm-newton: the pn-ccd camera,” *Astronomy & Astrophysics*, vol. 365, no. 1, pp. L18–L26, 2001.
- [30] J. Carter and A. Read, “The xmm-newton epic background and the production of background blank sky event files,” *Astronomy & Astrophysics*, vol. 464, no. 3, pp. 1155–1166, 2007.
- [31] D. C. Wells and E. W. Greisen, “Fits-a flexible image transport system,” in *Image Processing in Astronomy*, p. 445, 1979.
- [32] T.-P. Li and M. Wu, “Reconstruction of objects by direct demodulation,” *Astrophysics and Space Science*, vol. 215, no. 2, pp. 213–227, 1994.
- [33] J.-L. Starck and M. Pierre, “Structure detection in low intensity x-ray images,” *Astronomy and Astrophysics Supplement Series*, vol. 128, no. 2, pp. 397–407, 1998.

- [34] L. Faccioli, F. Pacaud, J.-L. Sauvageot, M. Pierre, L. Chiappetti, N. Clerc, R. Gastaud, E. Koulouridis, A. Le Brun, and A. Valotti, “The xxl survey-xxiv. the final detection pipeline,” *Astronomy & Astrophysics*, vol. 620, p. A9, 2018.
- [35] W. Xu, M. E. Ramos-Ceja, F. Pacaud, T. H. Reiprich, and T. Erben, “A new x-ray-selected sample of very extended galaxy groups from the rosat all-sky survey,” *Astronomy & Astrophysics*, vol. 619, p. A162, 2018.
- [36] C. Zhang, M. E. Ramos-Ceja, F. Pacaud, and T. H. Reiprich, “High-redshift galaxy groups as seen by athena/wfi,” *Astronomy & Astrophysics*, vol. 642, p. A17, 2020.
- [37] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [38] H. Chen, X. He, L. Qing, Y. Wu, C. Ren, R. E. Sheriff, and C. Zhu, “Real-world single image super-resolution: a brief review,” *Information Fusion*, vol. 79, pp. 124–145, 2022.
- [39] G. P. Garmire, M. W. Bautz, P. G. Ford, J. A. Nousek, and G. R. Ricker Jr, “Advanced ccd imaging spectrometer (acis) instrument on the chandra x-ray observatory,” in *X-Ray and Gamma-Ray Telescopes and Instruments for Astronomy*, vol. 4851, pp. 28–44, International Society for Optics and Photonics, 2003.
- [40] T. Dauser, S. Falkner, M. Lorenz, C. Kirsch, P. Peille, E. Cucchetti, C. Schmid, T. Brand, M. Oertel, R. Smith, *et al.*, “Sixte: a generic x-ray instrument simulation toolkit,” *Astronomy & Astrophysics*, vol. 630, p. A66, 2019.
- [41] C. Gabriel, A. I. Ibaibarriaga, and J. Hoar, “Scisim: the xmm-newton x-ray observatory data simulator,” in *UV, X-Ray, and Gamma-Ray Space Instrumentation for Astronomy XIV*, vol. 5898, p. 58981O, International Society for Optics and Photonics, 2005.
- [42] C. Schmid, R. Smith, and J. Wilms, “Simput-a file format for simulation input,” [heawww.harvard.edu/heasarc/formats/simput-1.0.0.pdf](http://hea-www.harvard.edu/heasarc/formats/simput-1.0.0.pdf), 2013.
- [43] D. Nelson, A. Pillepich, V. Springel, R. Pakmor, R. Weinberger, S. Genel, P. Torrey, M. Vogelsberger, F. Marinacci, and L. Hernquist, “First results from the tng50 simulation: galactic outflows driven by supernovae and black hole feedback,” *Monthly Notices of the Royal Astronomical Society*, vol. 490, no. 3, pp. 3234–3261, 2019.

- [44] A. Pillepich, D. Nelson, V. Springel, R. Pakmor, P. Torrey, R. Weinberger, M. Vogelsberger, F. Marinacci, S. Genel, A. van der Wel, *et al.*, “First results from the tng50 simulation: the evolution of stellar and gaseous discs across cosmic time,” *Monthly Notices of the Royal Astronomical Society*, vol. 490, no. 3, pp. 3196–3233, 2019.
- [45] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist, M. Vogelsberger, S. Genel, P. Torrey, F. Marinacci, *et al.*, “First results from the illustrisng simulations: matter and galaxy clustering,” *Monthly Notices of the Royal Astronomical Society*, vol. 475, no. 1, pp. 676–698, 2018.
- [46] F. Marinacci, M. Vogelsberger, R. Pakmor, P. Torrey, V. Springel, L. Hernquist, D. Nelson, R. Weinberger, A. Pillepich, J. Naiman, *et al.*, “First results from the illustrisng simulations: radio haloes and magnetic fields,” *Monthly Notices of the Royal Astronomical Society*, vol. 480, no. 4, pp. 5113–5139, 2018.
- [47] D. Nelson, A. Pillepich, V. Springel, R. Weinberger, L. Hernquist, R. Pakmor, S. Genel, P. Torrey, M. Vogelsberger, G. Kauffmann, *et al.*, “First results from the illustrisng simulations: the galaxy colour bimodality,” *Monthly Notices of the Royal Astronomical Society*, vol. 475, no. 1, pp. 624–647, 2018.
- [48] J. P. Naiman, A. Pillepich, V. Springel, E. Ramirez-Ruiz, P. Torrey, M. Vogelsberger, R. Pakmor, D. Nelson, F. Marinacci, L. Hernquist, *et al.*, “First results from the illustrisng simulations: A tale of two elements—chemical evolution of magnesium and europium,” *Monthly Notices of the Royal Astronomical Society*, vol. 477, no. 1, pp. 1206–1224, 2018.
- [49] A. Pillepich, D. Nelson, L. Hernquist, V. Springel, R. Pakmor, P. Torrey, R. Weinberger, S. Genel, J. P. Naiman, F. Marinacci, *et al.*, “First results from the illustrisng simulations: the stellar mass content of groups and clusters of galaxies,” *Monthly Notices of the Royal Astronomical Society*, vol. 475, no. 1, pp. 648–675, 2018.
- [50] K. Arnaud, “Xspec: The first ten years,” in *Astronomical Data Analysis Software and Systems V*, vol. 101, p. 17, 1996.
- [51] R. Gilli, A. Comastri, and G. Hasinger, “The synthesis of the cosmic x-ray background in the chandra and xmm-newton era,” *Astronomy & Astrophysics*, vol. 463, no. 1, pp. 79–96, 2007.
- [52] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, “Densenet: Implementing efficient convnet descriptor pyramids,” *arXiv preprint arXiv:1404.1869*, 2014.
- [53] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-time single image and video super-resolution using

an efficient sub-pixel convolutional neural network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1874–1883, 2016.

- [54] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image restoration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2480–2495, 2020.
- [55] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [56] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in *The Thirtieth Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2, pp. 1398–1402, Ieee, 2003.
- [57] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, “Loss functions for image restoration with neural networks,” *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.
- [58] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “Fsim: A feature similarity index for image quality assessment,” *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [59] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, “A haar wavelet-based perceptual similarity index for image quality assessment,” *Signal Processing: Image Communication*, vol. 61, pp. 33–43, 2018.
- [60] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [61] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” in *International conference on artificial neural networks*, pp. 270–279, Springer, 2018.

Appendix A

Appendix

A.1 Real XMM Dataset Generation

The real XMM-Newton dataset was created using the following pipeline:

- We start off with the full XMM-Newton archive, containing all the historical observations.
- First, we select all observations of at least 20ks observation time using either full-frame or extended full-frame mode. These modes use all the 12 EPIC-pn sensors. Smaller modes are usually used for brighter objects, which are not the focus of our research.
- We download all the eventlists of an observation. Usually, one observation is split into different recordings. We pick the eventlist with the longest on-time. The on-time is the planned observation time without unexpected interruptions. For example, the presence of high solar flares forces the instruments to temporarily shut down to prevent damage.
- We clean the eventlist by removing all the events outside the GTI (good time interval) determined by solar flares. Specifically, $\text{PI} > 300$ and pattern $<= 4$.
- We filter out bad events caused by other reasons ($\text{FLAG} == 0$), such as cosmic rays, counts close to CCD gaps, bad pixels, etcetera. We also exclude the values close to the readout size of every CDD ($\text{RAWY} > 12$) since these are usually not reliable.
- We now have our clean eventlist for this observation.
- We convert this eventlist into smaller eventlists of different exposure times with increments of 10 ks. The biggest exposure time depends on the exposure time of the cleaned eventlist. I.e. if we have a clean exposure of 40 ks,

we will generate 4x10 ks images, 2x20 ks images, 1x30 ks, and 1x40 ks images. The images with multiple exposure times make the dataset more flexible to use. It also enables us to train a de-noising model with low and high exposure image pairs.

- We also keep the readout of the corners that are blocked by the instrument. These corners cannot be hit with photons originating from the sky; all the counts, therefore, are background noise. We made this choice since the SIXTE simulations also contain this part, and therefore we can validate if the noise properties are similar.
- Finally, we create images from these eventlists. This is done by counting all the events for a specific CCD pixel. The count for each pixel can be visualized as the intensity, creating an image. We use the default binsize of 80 (4"/pix). The final image is saved in the *fits* format [31]. The fits file format is a lossless (image) file format used often in astronomy.

A.2 Simulation Setup Details

A.2.1 Boresight Determination

Note that the optical axis (also called boresight) is not exactly in the middle of the image but is slightly offset. This offset also changed over time. We used the information from the latest calibration file: XMM_MISCDATA_0022.CCF to determine the boresight of the simulations. This boresight is also important for the vignetting and the PSF, since these depend on the boresight, and not on the raw coordinates.

A.2.2 PSF

The PSF (point spread function) of the *XMM-Newton* is not constant as is explained in section 2.1.3. In SIXTE, there are two PSF implementations: Using a single PSF for the whole image or setting separate PSFs for certain sections. These sections are radially distributed, centering around the boresight using a polar coordinate system. For every X-ray photon of specific energy entering the simulated telescope, SIXTE will then use the closest given PSF for that specific energy and location. The PSF distributions that SIXTE uses need to be provided as images. During development, this created a problem since providing many PSF images, which make the simulation more realistic, resulted in very high memory use, limiting the number of simulations we could run in parallel.

We decided to use three different energy levels: 0.5, 1.0 and 2.0 KeV to optimize this. Use a ϕ degree interval of 4 degrees, and $\theta = 0, 210, 420, 600, 720, 900, 1200$ arcsec. Resulting in 630 unique PSF images for every energy level. The PSF image resolution was set to 120x120; this is just big enough to cover the most stretched PSF at the edge of the sensor. The PSF images were created

using the *psfgen* program, which is part of the official XMM-Newton Science Analysis System (XMM-SAS). To increase the simulation’s spatial resolution, we have to decrease the PSF size. However, we do want to keep the same PSF distortion shape. Therefore, we decreased the size of the original PSF images by the resolution multiplier.

A.3 Image Quality Metrics

In this section the implementation of the image quality metrics used in this thesis are defined. These metrics are used both for evaluation and loss functions.

A.3.1 L1

Given a predicted image \hat{y} and its target image y . The L1 loss used in this thesis is defined as:

$$L_1 = \frac{1}{N} \sum_{i=0}^N |y_i - \hat{y}_i| \quad (\text{A.1})$$

A.3.2 Poisson

Given a predicted image \hat{y} and its target image y . The Poisson loss used in this thesis is defined as:

$$\text{Poisson} = \frac{1}{N} \sum_{i=0}^N (\hat{y}_i - y_i \log(\hat{y}_i)) \quad (\text{A.2})$$

A.3.3 PSNR

Peak Signal-to-Noise Ratio (PSNR). PSNR measures the ratio between the maximum signal power and the power of the noise. Since the signals can have a wide dynamic range the PSNR is expressed logarithmic using the decibel scale. We use the PSNR formulation via the mean squared error (MSE). Given a (noisy) predicted image \hat{y} and its target image y , MSE and PSNR is defined as:

$$\text{MSE} = \frac{1}{N} \sum_{i=0}^N (y_i - \hat{y}_i)^2 \quad (\text{A.3})$$

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\max_y^2}{\text{MSE}} \right) \quad (\text{A.4})$$

Where \max_y is the maximum value of image y .

A.3.4 SSIM

Structural Similarity Index (SSIM) [55] has multiple parameters that can be tuned. In this research we are interested in the quality of the generated extended sources. The SSIM parameters control the influence of the small details and size of the structures on the final score. The final SSIM score is calculated based on the the SSIM quality map of an image. We can visualise this SSIM quality map to see the influence of these parameters. The final SSIM parameters used are:

- Data Range = The max clipping value of our input images at 200 times the mean background rate
- Window Size = 13
- Sigma = 2.5
- K1 = 0.01
- K2 = 0.05

A.3.5 MS_SSIM

Multi-Scale Structural Similarity Index (MS_SSIM) [56] calculates SSIM at different scales (window sizes). We used the same parameters as for SSIM with the weight for each scale being [0.0448, 0.2856, 0.3001, 0.2363, 0.1333].

A.3.6 FSIM and HaarPSI

The FSIM [58] and HaarPSI [59] metrics where used in the single channel mode with their default parameters.

A.4 Loss Function Normalization

To meaningfully combine loss functions, we need to normalize them since the values of different loss functions can differ by orders of magnitude. We normalize the loss based on the trial runs of the model trained with Poisson loss with the various data scaling functions and an untrained model. A model trained with a loss function different from the Poisson loss will generate different output images. These will have different loss values. However, the difference between a trained and untrained model with any of our loss functions will be huge. Therefore, the final loss metrics should be approximately on the same scale.

We aim to have the normalized loss value of the untrained model at 1.0 and the trained model at 0.0. We calculated the normalization with the following formula:

$$L_{normalized} = \alpha * L_{unnormlized} + \beta \quad (\text{A.5})$$

With:

$$\alpha = \frac{y_2 - y_1}{x_2 - x_1} \quad (\text{A.6})$$

$$\beta = y_1 - a \cdot x_1 \quad (\text{A.7})$$

Where y_1 is the target loss value for the untrained model (in our case $y_1 = 1$), y_2 is the target loss value for the trained model (in our case $y_2 = 0$), x_1 is the measured loss of the untrained model and x_2 is the measured loss of the trained model. We can now combine different loss functions by adding the normalized loss functions together since the loss functions are now on roughly the same scale.

Note that Equation A.5 transforms increasing metrics, such as PSNR and SSIM, to be decreasing. This is crucial if we combine them with decreasing loss metrics, such as L1, PSNR and Poisson.

A.5 IllustrisTNG Simput

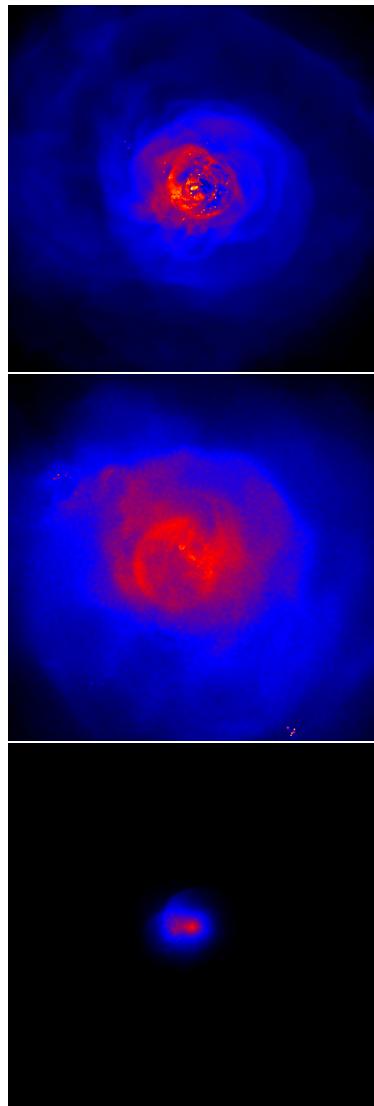


Figure A.1: Example IllustrisTNG simput images scaled with a logarithmic function. From *top* to *bottom*: TNG50-1 subhalo 382215, TNG100-1 subhalo 320012 and TNG300-1 subhalo 415728.

A.6 Simulated XMM-Newton Images

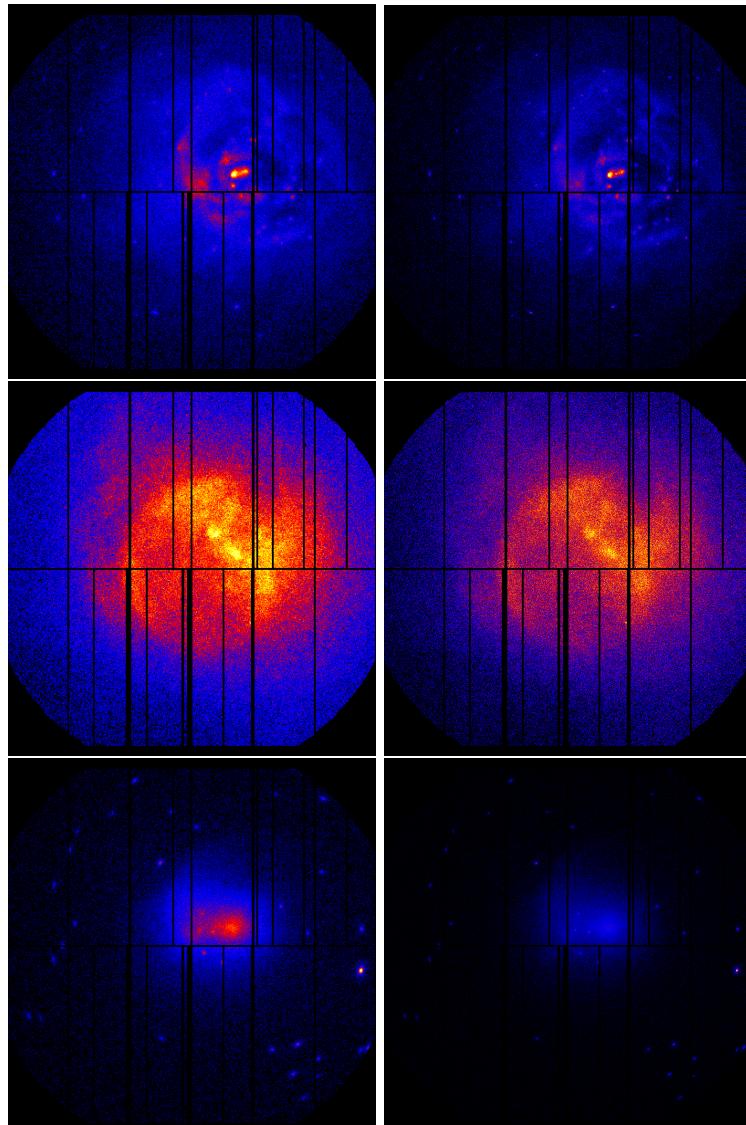


Figure A.2: Example images from the simulated dataset scaled with the square root function. The left column is at 1x resolution and the right column at 2x. From *top* to *bottom*: TNG50-1 subhalo 382215, TNG100-1 subhalo 320012 and TNG300-1 subhalo 415728 all at 100ks exposure time.

A.7 Full-Frame Super-resolution and De-noising

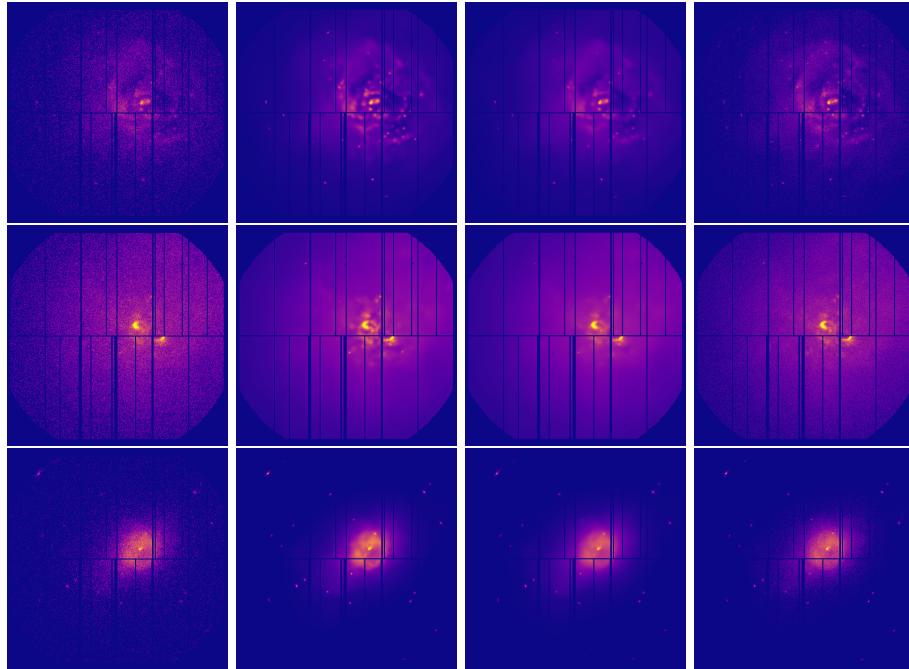


Figure A.3: Full-frame super-resolution and de-noising examples from the simulated XMM-Dataset scaled with the square root function. From *top* to *bottom*: TNG50 Subhalo 382215, TNG100 Subhalo 41583 and TNG300 Subhalo 296363. From *left* to *right*: Input image at 1x resolution with 20ks exposure, generated super-resolution image for 100ks with the SSIM loss model, generated super-resolution image for 100ks with the PSNR combined with MS_SSIM loss model and the label image at 100ks without background noise.

A.8 Full-Frame De-noised and Wavelet

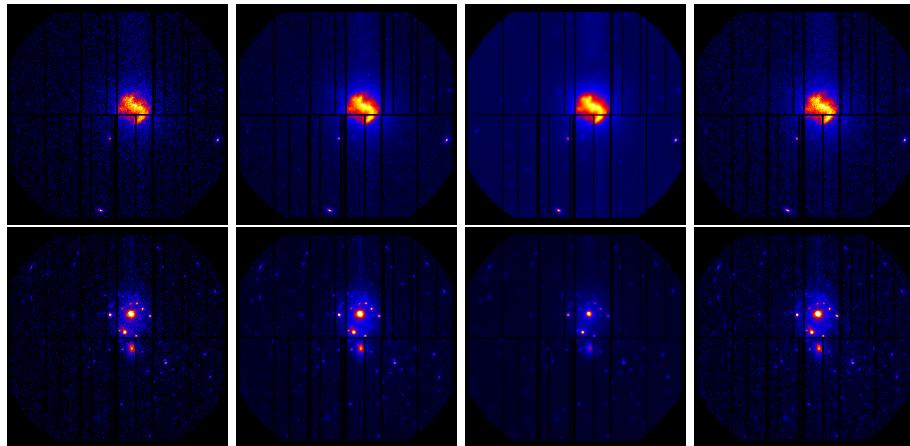


Figure A.4: Full-frame de-noised and wavelet transformed examples, W49B (top) and M51 (bottom), from the real XMM dataset scaled with the square root function. From *left* to *right*: Input image at 1x resolution with 20ks exposure, generated de-noised image for 50ks, wavelet transformed image and the target image at 50ks.

A.9 Bright Sources Limitation

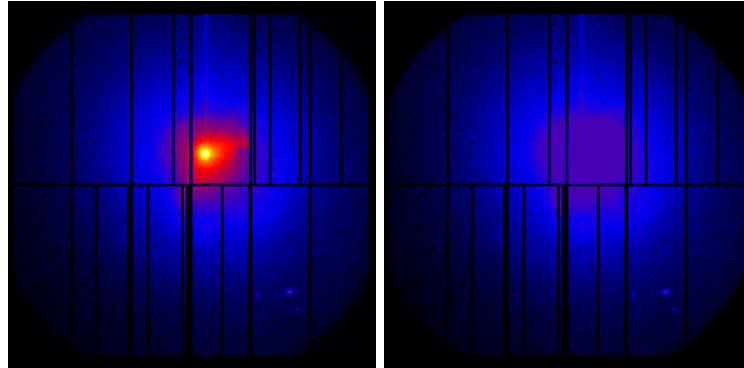


Figure A.5: *XMM-Newton* observations of the supergiant galaxy M87 (obsid 0200920101) scaled with the log function. The original observation at 20ks exposure (left) and the same observation clipped at 200 times the mean background rate (right).

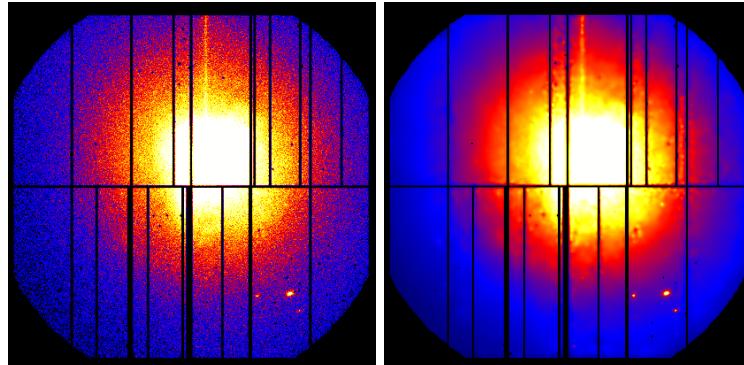


Figure A.6: Images of the supergiant galaxy M87 (obsid 0200920101) scaled with the sqrt function. Input image at 1x resolution with 20ks exposure clipped at 200 times the mean background rate (left) and the generated super-resolution and de-noised image (right).

To stabilise the training of the super-resolution model we clip pixel values to 200 times the mean background rate $\mu_B = 1.1168 \cdot 10^{-5} \text{counts/s}$. The majority of the extended features have X-ray counts below 200 times the mean background. However, this clipping can lead to the loss of detail for bright sources. This can be seen in Figure A.5, where the bright source M87 is shown without clipping (left) and with clipping (right). In the clipped image (right), all the pixels at the center are at the max clipping value and all detail is lost.

When this observation is enhanced with our SR and DN models, Figure A.6, we can see that they produce many artifacts. This can be explained by the fact that our simulator is developed for fainter sources. The models, therefore, never encountered such bright sources during training. The bright source also causes many out-of-time events, as can be seen by a straight line from the central source to the top of the image. Our simulator did not simulate these; consequently, the model never encountered them during training.